doi:10.1006/jmbi.2000.4012 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **301**, 915–945

JMB

1 Å Crystal Structures of *B*-DNA Reveal Sequence-specific Binding and Groove-specific Bending of DNA by Magnesium and Calcium

Thang Kien Chiu and Richard E. Dickerson*

Molecular Biology Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles CA 90095-1570, USA

**Corresponding author*

The 1 Å resolution X-ray crystal structures of Mg^{2+} and Ca^{2+} salts of the *B*-DNA decamers CCAACGTTGG and CCAGCGCTGG reveal sequence-specific binding of Mg^{2+} and Ca^{2+} to the major and minor grooves of DNA, as well as non-specific binding to backbone phosphate oxygen atoms. Minor groove binding involves H-bond interactions between cross-strand DNA base atoms of adjacent base-pairs and the cations' water ligands. In the major groove the cations' water ligands can interact through H-bonds with O and N atoms from either one base or adjacent bases, and in addition the softer Ca^{2+} can form polar covalent bonds bridging adjacent N7 and O6 atoms at GG bases. For reasons outlined earlier, localized monovalent cations are neither expected nor found.

Ultra-high atomic resolution gives an unprecedented view of hydration in both grooves of DNA, permits an analysis of individual anisotropic displacement parameters, and reveals up to 22 divalent cations per DNA duplex. Each DNA helix is quite anisotropic, and alternate conformations, with motion in the direction of opening and closing the minor groove, are observed for the sugar-phosphate backbone. Taking into consideration the variability of experimental parameters and crystal packing environments among these four helices, and 24 other Mg^{2+} and Ca^{2+} bound *B*-DNA structures, we conclude that sequence-specific and strand-specific binding of Mg^{2+} and Ca^{2+} to the major groove causes DNA bending by base-roll compression towards the major groove, while sequence-specific binding of Mg^{2+} and Ca^{2+} in the minor groove has a negligible effect on helix curvature. The minor groove opens and closes to accommodate Mg^{2+} and Ca^{2+} without the necessity for significant bending of the overall helix.

The program *Shelxdna* was written to facilitate refinement and analysis of X-ray crystal structures by *Shelxl-97* and to plot and analyze one or more *Curves* and *Freehelix* output files.

Keywords: divalent cations; DNA sequence-specific binding; groove-specific bending; anisotropy; *shelxdna*

Introduction

The influence of monovalent and divalent cations on the conformation and structure of duplex DNA has been studied widely by X-ray crystallography, NMR, gel electrophoresis, and molecular dynamics simulation (MDS). Monovalent Na^+ and K^+ and divalent Mg^{2+} and Ca^{2+} are the most abundant cations within cells, with basal concentrations of approximately 30 mM Na^+, 100 mM K^+, and 40 mM $[Mg^{2+} + Ca^{2+}]$ (Boynton *et al.*, 1982; Collins, 1997). In comparison, the molar concentration of pure water is ~55 M.

Abbreviations used: *CG*, CCAACGTTGG; *AG*, CCAGCGCTGG; *CI*, CCAACITTGG; *AT*, CGATATATCG; CGMg, *CG*:Mg^{2+}; CGCa, *CG*:Ca^{2+}; AGMg, *AG*:Mg^{2+}; AGCa, *AG*:Ca^{2+}; CIMg, *CI*:Mg^{2+}; *CI*Ca, *CI*:Ca^{2+}; ATMg, *AT*:Mg^{2+}; ATCa, *AT*:Ca^{2+}; VTCa, GGCGAATTCGCG:Ca^{2+}; VTMg, CGCGAATTCGCG:Mg^{2+}; ADPs, anisotropic displacement parameters; MDS, molecular dynamics simulations; MEP, molecular electrostatic potential.

E-mail address of the corresponding author: red@mbi.ucla.edu

0022-2836/00/040915–31 $35.00/0

Monovalent cations do not recognize much variation along the surface of a regular DNA duplex. The only discriminating features are differences in molecular electrostatic potential (MEP) and hydrogen bond environment between major and minor grooves. Monovalent cations themselves tend to be less solvated than divalent cations. For example, the modified Born equation for metal hydration (Draper & Misra, 1998), $\Delta G = -164\ Z^2/(r + 0.7)$, where Z and r are the cation charge and radius, gives a free energy of hydration that is five times more favorable for Mg^{2+} than for Na^{+}. Hence monovalent cations tend to interact with DNA electrostatically and non-specifically, without using hydrogen bonds from coordinated water ligands. Their interactions with *B*-DNA are ''loose and delocalized'' in the terminology of Braunlin (1995), and exhibit preferential but fractional binding to the minor groove of A-tracts as demonstrated by NMR quadrupole relaxation (Denisov & Halle, 2000), X-ray crystallography (Tereshko *et al*., 1999a,b), and MDS (Young *et al*., 1997; Feig & Pettitt, 1999). This weak and changeable monovalent cation binding therefore is unlikely to affect the average conformation of duplex DNA under physiological conditions (Yuan *et al*., 1992; Young *et al*., 1995, 1997; Young & Beveridge, 1998; Zacharias *et al*., 1982).

In marked contrast, divalent cations are widely known to affect the structure of duplex DNA. Mg^{2+}, Ca^{2+}, Zn^{2+}, Co^{2+}, Ba^{2+}, Mn^{2+}, and Cd^{2+} can induce *B*-to-*Z* transition of poly(d(G-C)) sequences (Zaccharias *et al*., 1982; van de Sande *et al*., 1982; Narasimhan & Bryan, 1984; Behe & Felsenfeld, 1981) and Co^{3+} can induce *B*-to-*A* transitions (Narasimhan & Bryan, 1984). At physiological concentrations the binding of divalent cations to DNA is both cation-dependent and sequence-dependent. From a sequence standpoint, specificity is contributed by both the local molecular electrostatic potential and the hydrogen bond environment. These H-bond interactions reflect the greater hydration properties of divalent cations over monovalent. The affinity of each divalent cation for DNA is a function of its particular degree of softness, hydration free energy and hydration geometry. For example, the soft transition metals Mn^{2+} and Co^{2+} have much higher affinity for the soft guanine N7 than do Mg^{2+} and Ca^{2+} (Zaccharias *et al*., 1982; van de Sande *et al*., 1982). Similarly, Mg^{2+} chelated to a bulky chlorophyll-a molecule can only interact with the DNA backbone phosphate oxygen (~60 %) and atoms of the wide major groove (~20 % to guanine N7, probably through a coordinated water molecule) (Neault & Tajmir-Riahi, 1999).

Because of their greater affinity and sequence specificity, the binding of divalent cations to DNA can result in DNA bending. Among the earliest evidence of this is the decrease of electrophoretic mobility of kinetoplast DNA fragments in the presence of Mg^{2+} (Diekmann & Wang, 1985). However, despite their observed sequence-dependent binding and bending of duplex DNA, the structural roles of divalent cations, especially Mg^{2+} and Ca^{2+}, have so far remained relatively unappreciated. Current attempts to establish sequence-to-structure relationships based on analysis of the available crystallographic data or energetic analysis of base stacking interactions (El Hassan & Calladine, 1995; Young *et al*., 1995; Hunter & Lu, 1997; Dickerson, 1998) have not considered structural contributions from bound cations.

At most, it has been recognized that Mg^{2+} and Ca^{2+} can influence the packing of duplex DNA in crystals. The *CI* decamer (Lipanov *et al*., 1993) and CGCGAATTCGCG dodecamer (Liu *et al*., 1998; Minasov *et al*., 1999) each crystallize in different space groups when grown with Ca^{2+} *versus* Mg^{2+}, whereas the *AT* decamer crystallizes in the same space group regardless of cation type (Table 1). With CICa, Ca^{2+} is localized in the minor groove at AA and TT steps, while in CIMg, Mg^{2+} is localized in the major groove between bases C9 and G10. With VTCa, a Ca^{2+} again is localized in the minor groove at GA and CG steps (Minasov *et al*., 1999) while in VTMg, a Mg^{2+} is localized in the major groove between base-pairs G2 and C3 (Tereshko *et al*., 1999a). Unfortunately, it is difficult to assess the contributions of base sequence or cation type in the *CI* and *VT* sequences because of inherent differences in crystal packing between Mg^{2+} and Ca^{2+} forms. Indeed, most cations in VTMg and VTCa bind to the sugar-phosphate backbone, supporting their roles in intermolecular crystal packing. In both ATMg and ATCa, one cation is observed in the minor groove between base-pairs T4-A17 and A5-T16. But it is possible that binding at other sites was not detected because of the relatively low resolution of these structures.

We have previously observed a cation-dependent deformation of DNA in the crystal structure of the crosslinked dodecamer CGCGAATTCGCG in which the N6 atoms of bases A6 and A18 are linked by a -CH_2-CH_2-S-S-CH_2-CH_2- tether (Chiu *et al*., 1999). The tether is long enough, and the connected adenine bases near enough, that no deformation of the helix results from crosslinking. Although the sequence is palindromic and end-for-end crystal packing environments are similar (see Figures 9 and 10 of Chiu *et al*., 1999), binding of a Mg^{2+} complex in the major groove between base-pairs 2-3, but not between base-pairs 10-11, leads to asymmetric kinking of the helix near base-pairs 3-5 in the direction of the major groove. This same asymmetric kinking was observed in the first *B*-DNA crystal structure, an un-crosslinked helix with the same sequence, but at the time was attributed wrongly to a bound spermine cation (Drew & Dickerson 1981).

To explore further the role of divalent cation binding on DNA bending, we have determined the crystal structures of the Mg^{2+} and Ca^{2+} salts of decamers CCAACGTTGG (the *CG* decamer) and CCAGCGCTGG (the *AG* decamer) at a resolution of 0.99 Å. These have been assigned PDB accession

numbers 1en3, 1en8, 1en9, and 1ene, respectively, or NDB accession numbers bd0033, bd0034, bd0035 and bd0036. All four are isomorphous, with identical crystal packing environments.

Because crystallization and data collection parameters among these four structures also are quite similar, any differences in cation binding and DNA structure can only be attributed either to base sequence or to cation type. To obtain a consensus picture of divalent cation binding to *B*-DNA in general, we also have compared with them the experimental parameters, cation binding preference, and bending behavior of 24 other crystalline *B*-DNA oligmers (Table 1). This database of 28 crystal structures spanning ten different crystal packing environments will be used to obtain a model for sequence-specific binding and groove-specific bending of *B*-DNA by divalent cations.

Results

Atomic resolution reveals multiple discrete conformations for DNA and cations

All four of the 0.99 Å structures, CGCa, CGMg, AGCa and AGMg, are isomorphous with the original 1.4 Å monoclinic C2 structure of sequence CCAACGTTGG (bdj019; Privé *et al*., 1991). That earlier structure was obtained with one-third as many reflections, which were modeled anisotropically with six tensor parameters. The much higher resolution and quality of the current data sets (Table 2A) allow the accurate modeling of discrete alternate conformations, individual ADPs for all heavy atoms, and hydrogen atoms riding at ideal positions, with final data-to-parameter ratios of around 6:1 (Table 2B). Luzzati analysis shows that these structures are roughly 2.5 times more accurate than bdj019 (coordinate errors of 0.06 *versus* 0.15 Å). The asymmetric unit for all seven sequences listed at the top of Table 1 consists of one half of the decamer duplex: either one of the two strands, or alternatively the first five base-pairs in from one end of the helix. The latter choice will be used in this paper; symmetry-equivalent cation and solvent atoms in the second half of the helix will be primed (′). As is customary, bases are numbered 1-10 in the first strand and 11-20 in the other. The crystallographic asymmetric unit, base-pairs C1·G5 through C16·G20, contains ten DNA bases with ca. 202 atoms, 93 to 126 solvent atoms (including alternative conformations), and four to six hydrated complexes of divalent cations. For reasons outlined by Chiu *et al*. (1999), localized monovalent cations are neither expected nor found.

The divalent cation complexes in these four structures, CGMg, CGCa, AGMg and AGCa, are shown in Tables 3 and 4 and stereo Figure 1, and are labeled by serial numbers 111 to 116. Complexes with the same number in different structures occupy corresponding if not identical positions. Water oxygen atoms around a given metal complex are denoted OW1 through OW6 (OW7 in the case of Ca) in an arbitrary manner, but once again similarly numbered atoms in different helices occupy analogous positions. Interatomic distances between the metals and their water ligands, and DNA oxygen and nitrogen atoms are listed in Table 3, and metal-ligand bond distances and thermal vibration parameters or *B*-values in Table 4.

Divalent cations cement helices together

The CGMg helix in its crystallographic environment is depicted in Figure 2, and the other three helices are very similar. Helices are stacked atop one another to simulate infinite columns, and columns are packed side-by-side. As was mentioned in connection with bdj019 (Prive *et al*., 1991), the stacked helices are continuous; the step from the last base-pair of one decamer to the first base-pair of the next along a given column is in no sense different from other base-pairs in the helix aside from the absence of connecting phosphates.

It is apparent from Figure 2 that lateral packing of helix columns would be impossible without charge neutralization of the phosphate backbone by cations. CGMg, CGCa, AGMg and AGCa have 6, 8, 12 and 14 cations per decamer duplex bound to the phosphate backbone, corresponding to fractional phosphate charge neutralization of 0.33, 0.44, 0.67 and 0.78, respectively. The ca. 20 % greater charge neutralization by Ca^{2+} than by Mg^{2+} in both sequences is undoubtedly real, since crystal packing, solution cation concentrations and data-collection protocols within each sequence are identical (see Discussion). In contrast, the roughly twofold increase in charge neutralization from *CG* to *AG* is most likely an experimental artifact, since both the cation concentration in the crystallizing solution and the measured I/σ signal in the X-ray data were greater for *AG* than for *CG*.

This increased affinity of Ca^{2+} over Mg^{2+} has been observed previously in solution studies (Eichhorn & Shin, 1968; Langlais *et al*., 1990; Korolev *et al*., 1999). It was not observed in three previous Mg^{2+}/Ca^{2+} crystal structure comparisons because of inequalities of data resolution between pairs of structures (Yuan *et al*., 1992; Lipanov *et al*., 1993; Minasov *et al*., 1999). The absence of backbone binding for the Yuan *et al*. pair was most likely an artifact of low data resolution. The data were better for Lipanov and Minasov, but the resolution again was sufficiently different between Mg^{2+} and Ca^{2+} forms to bias their results (see Discussion). Moreover, differences in crystal packing between non-isomorphous Mg^{2+} and Ca^{2+} forms could presumably influence the binding of cations to the backbone. The remaining 10-53 % of the cations required for complete charge neutralization in the *CG* and *AG* decamers (maximum of ∼88 % for divalent cations; Li *et al*., 1998) are disordered and presumably mobile in the interstitial medium. In solution, since these interactions with the DNA are

Table 1. *B*-DNA structures with bound divalent cations

NDB	Name	Sequence	Cation	M/P	V_m	Volume (Å^3)	GC	M1 (a.u.)	M2 (m)	M3 (M)	M4 (sp)	M5 (all)	M23 (gr)	M4b (po)	M5b (bp)	Res (Å)	DC	Sp.Grp.	Reference
bd0033	CGMg	CCAACGTTGG	Mg^{2+}	2.0	2.08	25138	60	4	4	4	6	12	0.80	0.33	1.20	0.99	SC	C2	This work
bd0034	CGCa	CCAACGTTGG	Ca^{2+}	2.0	2.02	24762	60	4	4	2	8	13	0.60	0.44	1.30	0.99	SC	C2	This work
bd0035	AGMg	CCAGCGCTGG	Mg^{2+}	10.5	2.06	24864	80	6	4	6	12	22	1.00	0.67	2.20	0.99	SC	C2	This work
bd0036	AGCa	CCAGCGCTGG	Ca^{2+}	20.0	2.06	24834	80	4	4	4	14	18	0.80	0.78	1.80	0.99	SC	C2	This work
bdj019		CCAACGTTGG	Mg^{2+}	12.8	2.15	25978	60	4	4	4	4	10	0.80	0.22	1.00	1.40	K5°	C2	Privé *et al*. (1991)
bdj008		CCAAGATTGG	Mg^{2+}	13.0	2.10	25556	50	3	2	2	6	10	0.40	0.33	1.00	1.30	K5°	C2	Privé *et al*. (1987)
bdjb44	CICa	CCAACITTGG	Ca^{2+}	2.3	2.13	25568	60	4	2	0	2	14	0.20	0.11	1.40	1.30	K-25°	C2	Lipanov *et al*. (1993)
bdl042		**CGT**AGATCT**ACG**	Mg^{2+}	2.7	2.00	58077	17	1	0	0	1	1	0.00	0.05	0.08	2.25	K15°	C2	Leonard *et al*. (1993)
bdj060		CTCT**CG**AGAG	Ca^{2+}	4.1	2.05	49493	40	6	3	2	9	15	0.60	0.41	1.50	1.70	KC	C2	Goodsell *et al*. (1995)
bdjb43	CIMg	CCAACITTGG	Mg^{2+}	2.3	2.51	90628	60	1	0	1	1	2	0.10	0.06	0.20	2.20	K-25°	$P3_221$	Lipanov *et al*. (1993)
bdjb48		CGATCGA^{M}TCG	Mg^{2+}	22.8	2.61	94845	60	2	0	2	1	3	0.20	0.06	0.30	2.00	K0°	$P3_221$	Baikalov *et al*. (1993)
bdj052		CCAAGCTTGG	Ca^{2+}	19.1	2.31	83741	60	2	2	0	4	6	0.20	0.22	0.60	1.90	K5°	P6	Grzeskowiak *et al*. (1993)
bd0018		**GC**GAATTC**GCG**	Ca^{2+}	0.5-1.8	2.14	128941	33	6	1	0	10	11	0.17	0.45	1.10	1.30	SC	H3	Minasov *et al*. (1999)
bd0019	VTCa	**GGC**GAATTC**GCG**	Ca^{2+}	0.5-1.8	2.30	152007	33	4	2	0	6	8	0.33	0.27	0.80	1.70	SC	R3	Minasov *et al*. (1999)
bd0007	VTMg	**CGC**GAAT^{F}TC**GCG**	Mg^{2+}	1.9	2.19	63989	33	5	0	1	12	13	0.17	0.55	1.08	1.10	SC	$P2_12_12_1$	Tereshko *et al*. (1999a)
bd0012		**CGC**GAAT^{F}TC**GCG**/Rb	Mg^{2+}	1.0	2.25	65677	33	3	0	1	3	7	0.17	0.14	0.58	1.20	SC	$P2_12_12_1$	Tereshko *et al*. (1999b)
bdl084		**CGC**GAATTC**GCG**/Spm	Mg^{2+}	0.5	2.29	66488	33	1	0	1	1	2	0.17	0.05	0.17	1.40	KC	$P2_12_12_1$	Shiu *et al*. (1998a)
bd0008		**CGC**GAA^{RS}TTC**GCG**	Mg^{2+}	4.5	2.13	64597	33	3	0	1	6	6	0.17	0.27	0.50	1.43	KC	$P2_12_12_1$	Chiu *et al*. (1999)
bd0013		**CGC**GAAT^{F}TC**GCG**/Rb	Mg^{2+}	0.2	2.24	65492	33	2	0	1	4	5	0.17	0.18	0.42	1.50	KC	$P2_12_12_1$	Tereshko *et al*. (1999b)
bd0005		**CGC**GAATTC**GCG**/Spm	Mg^{2+}	0.9	2.31	67127	33	1	0	1	1	2	0.17	0.05	0.17	1.50	MC	$P2_12_12_1$	Shiu *et al*. (1998b)
bdlb84		**CGC**GAAT^{F}T^{F}C**GCG**	Mg^{2+}	0.5	2.22	65142	33	1	0	1	1	2	0.17	0.05	0.17	1.55	KC	$P2_12_12_1$	Berger *et al*. (1998)
bdlb85		**CGC**GAAT^{F}TC**GCG**	Mg^{2+}	0.5	2.25	65725	33	1	0	1	1	2	0.17	0.05	0.17	1.55	KC	$P2_12_12_1$	Berger *et al*. (1998)
bdls80		**CGC**GAAT^{CO}T^{CO}C**GCG**	Mg^{2+}	0.7	2.26	66618	33	1	0	1	1	2	0.17	0.05	0.17	2.14	K?	$P2_12_12_1$	Portmann *et al*. (1997)
bdj036	ATCa	CGATATATCG	Ca^{2+}	23.6	2.17	52373	40	1	1	0	1	2	0.10	0.06	0.20	1.70	K0°	$P2_12_12_1$	Yuan *et al*. (1992)
bdj037	ATMg	CGATATATCG	Mg^{2+}	2.5	2.13	51489	40	1	1	0	1	2	0.10	0.06	0.20	2.00	K0°	$P2_12_12_1$	Yuan *et al*. (1992)
bdj025		CGATCGATCG	Mg^{2+}	22.8	2.13	51375	60	2	1	1	1	3	0.10	0.06	0.30	1.50	K0°	$P2_12_12_1$	Grzeskowiak *et al*. (1991)
bdj031		CGATTAATCG	Mg^{2+}	22.9	2.07	49911	40	1	1	0	1	2	0.10	0.06	0.20	1.50	K0°	$P2_12_12_1$	Quintana *et al*. (1992)
bd0006		**GG**CCAATTGG	Mg^{2+}	5.8	2.04	50020	50	2	2	0	5	6	0.50	0.28	0.63	1.15	SC	$P2_12_12_1$	Vlieghe *et al*. (1999)

NDB, Nucleic Acid Data Base identification number. Name, Short reference name for frequently-mentioned helices. Sequence, Base sequence with bold face in regions where minor groove is obstructed by crystal packing. M, methyl; F, fluoro; RS, disulfide crosslink; CO = carbonyl. M/P, Number of cations in crystallization buffer per phosphate group of helices. V_m, Matthews number, or volume per Dalton in the crystal. Volume, Volume of unit cell in Å^3. GC, Percent GC among base-pairs having an unobstructed minor groove. M1 (a.u.), Cations per asymmetric unit. M2 (m), Minor groove cations per helix duplex. M3 (M), Major groove cations per duplex. M4 (sp), Cations interacting with sugar/phosphate backbone per duplex. M5 (all), Total cations per duplex. (M5 is smaller than M2 + M3 + M4 if a given cation interacts with both a groove and the helix backbone.) M23 (gr), Cations interacting with either groove, per base-pair having unobstructed grooves. M4b (po), Cations interacting with phosphate backbone, per phosphate group. M5b (bp), Total cations per base-pair. Res, Resolution of the X-ray data in Å. DC, Data collection method: S, synchrotron; K, copper target; M, molybdenum target; C, cryogenic (100 K); −5 Å, temperature in degrees Celsius. Sp.Gp., Crystallographic space group. Horizontal lines separate sets of isomorphous structures.

non-covalent, non-specific, short-lived (<1 ns), and equivalent (i.e. no phosphate oxygen atom is neutralized more than any other), they are unlikely to induce DNA bending. In contrast, permanent covalent attachment of positive partial charges to the phosphate oxygen atoms does indeed result in bending in the direction of the charge (Strauss-Soukup *et al*., 1997).

As shown in Figure 1, most of the interactions with the DNA backbone involve H-bonds between the cations' water ligands and atoms O1P and O2P. In some cases a hydrated cation bridges the O1P and O2P atoms of the same phosphate group, employing one or two of its water ligands. In other cases, multiple interactions with adjacent residues are observed. Mg^{2+} and Ca^{2+} also can interact ionically with the DNA backbone: one or more of their water ligands are replaced by an O1P and/or O2P atom(s) from at least one DNA helix. These ionic interactions were observed for Ca^{2+} in CGCa, VTCa, bdj060 and bdj052, and for Mg^{2+} in bdj008, bdjb48, VTMg and bd0008.

Divalent cation binding is sequence-specific

As mentioned in the Introduction, solution studies have long established that divalent cation binding to DNA is sequence-specific. As shown in Figure 1 and Table 3, divalent cations do not simply bind passively to the sugar-phosphate backbone. Because of their much greater hydration properties (by comparison with monovalent cations), they can utilize their bound water molecules to sense fine differences in hydrogen-bond environment in the grooves of DNA. We tested this sequence-discriminating property specifically by changing base 4 from A to G. The specific interactions will be discussed below but the overall differences in cation binding are summarized in Figure 3. In Figure 3, only cation interactions with DNA base atoms are given; non-specific interactions with the sugar-phosphate backbone are omitted.

As will be discussed in the Discussion, we note that interpretations of cation binding specificity in different crystal structures determined under different experimental conditions must be taken with caution, as the number of cations actually observed in the structure can be influenced by crystal packing, cation concentration in the crystallization droplet, data quality, and data resolution. This experimental bias is minimal in the four structures reported here because of the similarity of experimental conditions. Hence, under identical crystal packing environments, cation binding sites changed when the base sequence was changed.

For the major groove metal complex 114, Mg^{2+} but not Ca^{2+} binds bases G16-T17 in *CG* and both cations bind bases A3-G4 in *AG*. In *CG*, no cations are observed for the A3-A4 base step because the two contiguous N6 amino groups create a positive MEP, which repels the positively charged cations. This is also the case in the isomorphous CICa and bdj008, which were solved by different investigators to a high enough resolution (1.3 Å) that experimental bias is minimal. When the amino groups of residues 4 and 14 are replaced by an oxygen, as in the *AG* sequences, the MEP is negative and hence selective for cation binding. No cations are observed at the A4-G5 base steps in bdj008 probably because of the positive MEP contributed by the flanking 5′ A3. Note that had cation binding been non-selective and had crystal-packing forces been the primary determinant of cation binding to bases, the binding sites in these isomorphous structures would have remained the same. In addition, at the terminal two base-pairs, where the sequence is conserved among *CG*, *AG*, and CICa sequences, cation binding is the preserved. The mode of binding differs between Mg^{2+} and Ca^{2+}, but the locus of binding is constant.

Binding of divalent cations in the minor groove is influenced by both the local MEP and groove width, both of which are sequence-dependent. In the minor groove of *CG*, complex Mg^{2+} 112 binds base-pairs A4-C5/G16-T17 while the larger Ca^{2+} binds base-pairs A3-C5/G16-T18; this is also the case in CICa. When residue 4/17 is changed to G/C from A/T, additional cations are localized so that both Mg^{2+} and Ca^{2+} bind base-pairs A3-C5/G16-T18.

Major groove binding is strand-specific (complexes 113 and 114)

Hydrated cations 111 and 112 within the minor groove are centered in the groove, whose narrowness precludes association with one backbone chain or the other (Figure 1). In contrast, cations 113 and 114 in the broader major groove bind to one side of that groove. Mg113 binds to adjacent guanine bases G19 and G20, which Pullman & Pullman (1981) characterize as the most electrostatically favorable locus within the major groove. Two distinct conformations for cation Mg113 are observed in CGMg (differentiated in Figure 1 by shades of green), *versus* one conformation in bdj019. In AGMg, only one conformer of Mg113 is found, but a nearby Mg116 has 40% occupancy. Two of the water molecules coordinated to Mg^{2+} in CGMg donate H-bonds to O6 and N7 atoms of the 5′ guanine, while a third donates an H-bond to a nearby phosphate oxygen. By contrast, in AGMg a water ligand accepts H-bonds from the amino nitrogen atoms of C1 and C11.

Because Ca^{2+} is both a softer cation and has a less favorable free energy of hydration than Mg^{2+}, it prefers to bind directly GG base atoms. Adjacent guanine bases G19 and G20 are locked together by polar-covalent bonds from Ca^{2+} to the N7 of G19 and the O6 of G20, as well as hydrogen bonds between its water ligands, nearby phosphate oxygen atoms, and the otherwise unused O6 of G19 and the N7 of G20. The attached cytosine bases, C1 and C2, are unliganded and free to propeller twist.

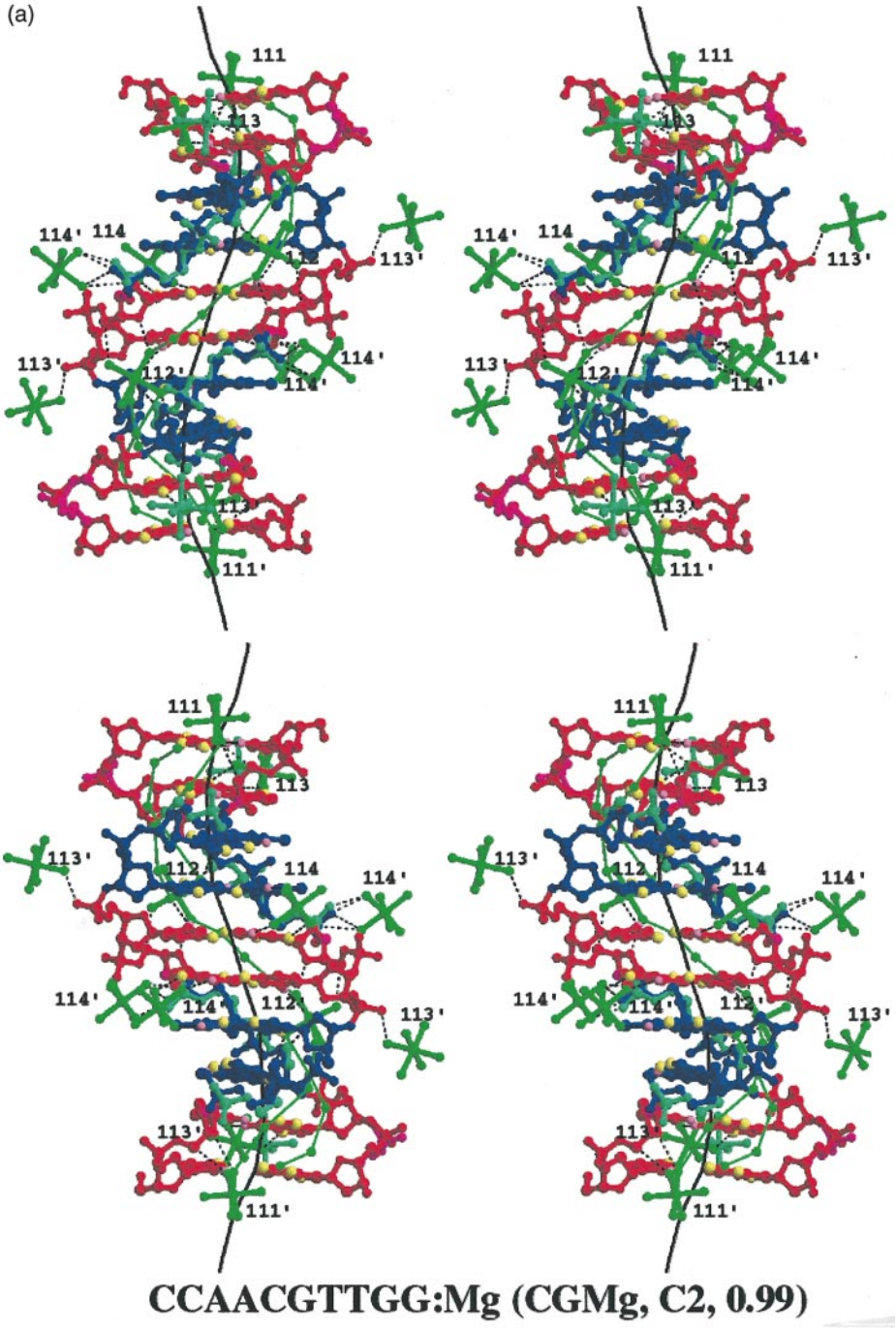

Figure 1 (*legend shown on page 923*)

An initial examination of the isomorphous structure CICa showed an apparent anomaly: a water molecule, and not the hydrated calcium complex Ca113, is present at the GG bases in the published structure. However, visual inspection of the final $F_o - F_c$ map recalculated in our laboratory and contoured at 1.5 s showed additional positive density at the presumed calcium position. Using the criteria elaborated in the Materials and Methods, this solvent peak, water residue 30, and arsenic residue 11, were reassigned as calcium complexes 113, 114, and 111, respectively. Substitution of three calcium complexes for three water molecules, followed by refinement, decreased the crystallographic *R* factor from 15.06 % to 11.25 %. Note that the absence of sodium cacodylate (which contains trace amounts of arsenic) in the crystallization buffers of CGCa and AGCa supports the reassignment of the arsenic residue as Ca111.

Like 113, complex 114 recognizes a single strand within the major groove, binding to G16-T17 in CGMg and A3-G4 in AGMg and AGCa. In all

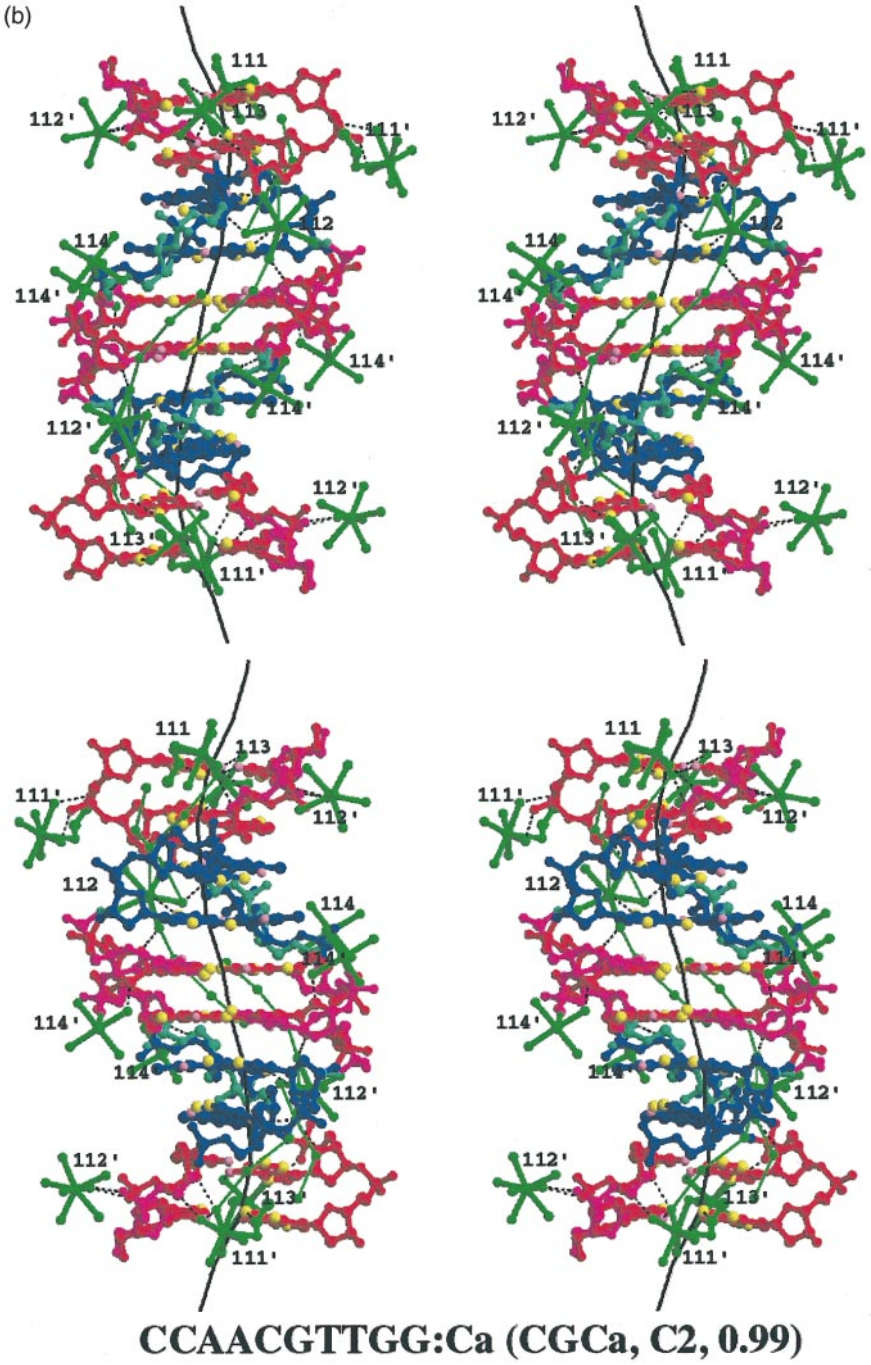

Figure 1 (*legend shown on page 923*)

three cases, the cations' water ligands donate H-bonds to atoms of both bases and to a nearby phosphate group. In CGCa, Ca114 does not recognize bases 16-17 at all, but instead interacts exclusively with phosphate backbone atoms. This is also true in CICa (note that guanine and inosine are identical within the major groove). Although the same backbone atoms are involved, interactions occur *via* ionic bonds with Ca^{2+} in CGCa, but involve H-bonds with water ligands of Ca^{2+} in CICa. This difference further supports the non-specific nature of cation-backbone binding. The direct consequence of four ionic bonds between neighboring helices in CGCa (at residues 6, 7, 16, and 17) is that the central four base-pair region of the minor groove is pulled open by as much as 2 Å. One could argue for a reversal of cause and effect, claiming that these ionic interactions are made possible only because of the wider minor groove. But this is unlikely, since the groove is nar-

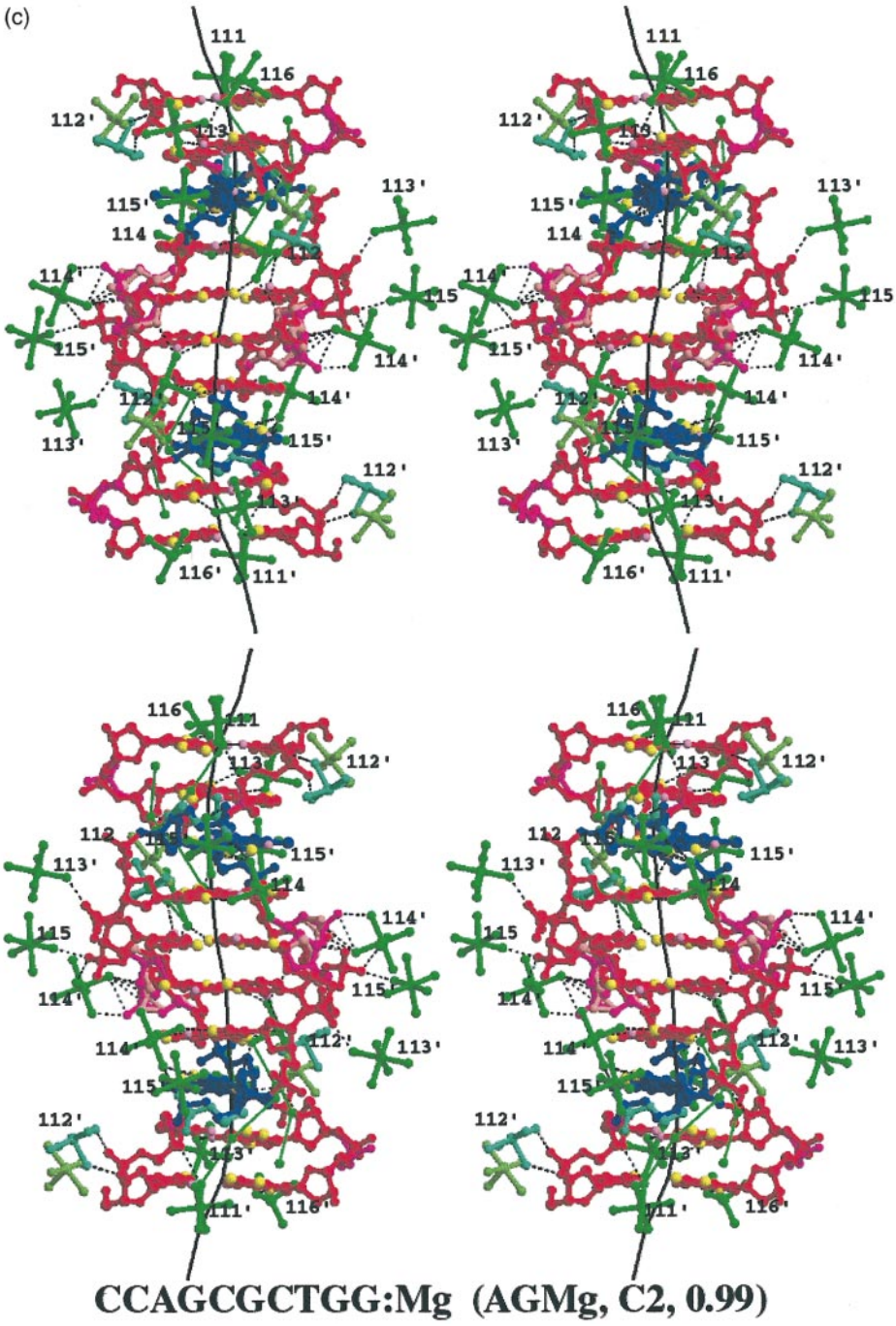

Figure 1 (*legend shown on page 923*)

rower in CICa, where cations merely H-bond with the DNA backbone. In addition, the introduction of a single inosine has no effect on DNA structure (Shatzky-Schwartz *et al*., 1997).

Minor groove binding involves cross-strand interlocking of base steps (complexes 112 and 111)

Because the minor groove of *B*-DNA is much narrower than its major groove, cation binding to the minor groove usually involves interactions between the cation's water ligands and cross-strand base and O4′ atoms of adjacent base-pairs. A fully hydrated cation usually sits at the center of the groove while each of its water ligands at maximum can donate two and accept one H-bond to base and backbone atoms. In addition, the regular octahedral geometry of magnesium gives an effective ion size of ca. 5.4 Å while the irregular geometry of calcium gives an effective ion size of ca. 5 to 7 Å. Hence, in *CG*, Ca112 sits between base-pairs 3

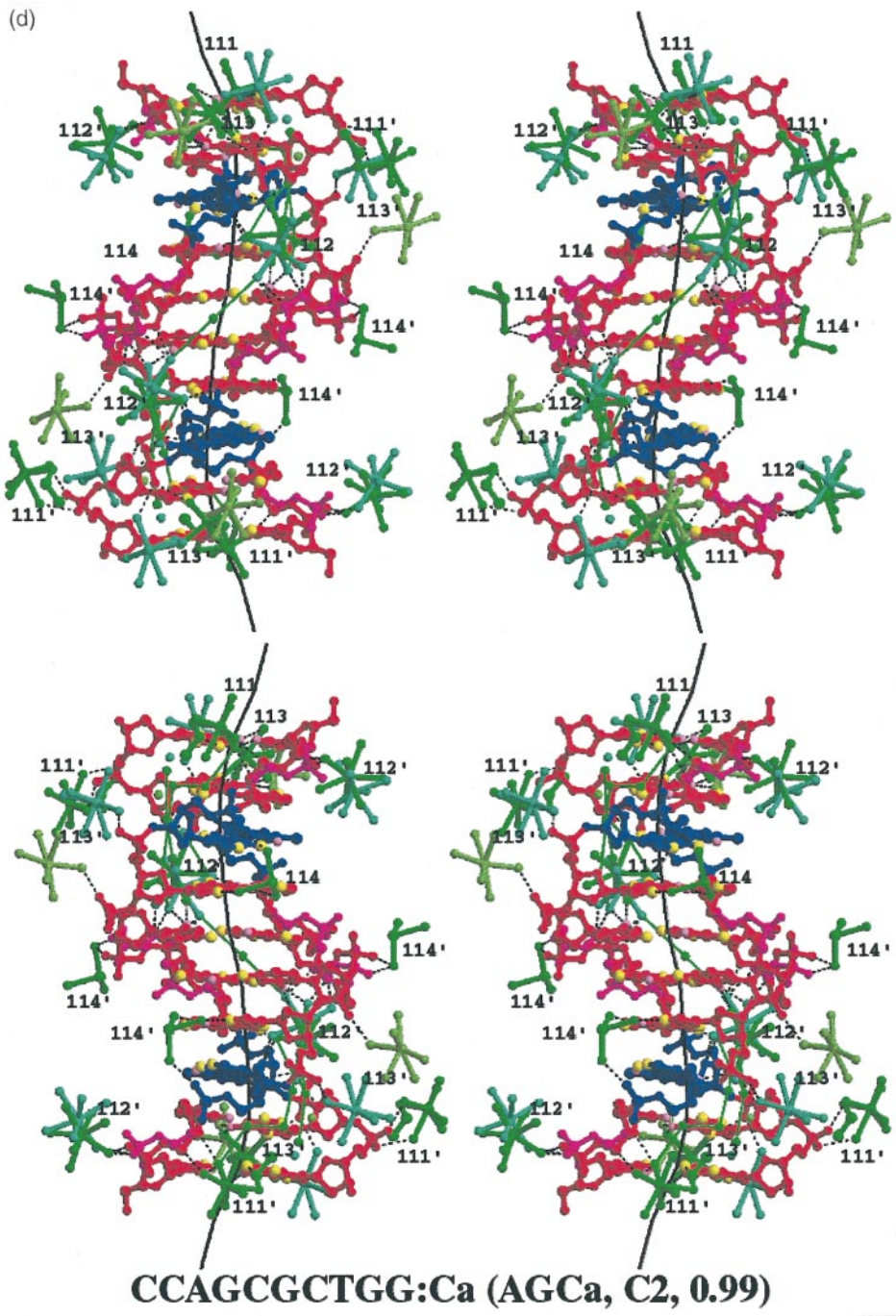

Figure 1. Stereograms of (a) CGMg, (b) CGCa, (c) AGMg and (d) AGCa. CG base-pairs are in red or magenta, and AT base-pairs are in blue and cyan, with magenta and cyan denoting alternative conformations. N and O atoms of each base-pair are yellow and pink, respectively. Ribbons of hydration along the minor groove floor are depicted by green lines connecting small green atoms. For clarity, hydrogen bonds between the water molecules of the ribbons and DNA atoms are omitted. Base-pair C1·G20 is at the top of each helix, and G10·C11 is at the bottom. All cations lying within hydrogen-bonding distance of the central helix are shown in green, with cyan and lime green denoting alternative conformations. Cations 111 and 112 sit within the minor groove, 111 at the top of the helix and 112 five base-pairs down. Cations 113 and 114 occupy the major groove, 113 near the top and 114 to one side about halfway down. Gently curved axes through the center of the helix are drawn with CURVES (Lavery & Sklenar, 1988). See Figures 3 and 11 also.

and 4 while the smaller Mg112 sits between base-pairs 4 and 5 in a narrower region of the groove. When the solution cation concentration is increased five to ten times (as from *CG* and *CI* to *AG*), additional conformations appear: one between base-pairs 4 and 5 in AGCa and two between base-

Table 2. Crystallization, data collection and refinement statistics of 0.99 Å structures

	CGCa	CGMg	AGCa	AGMg
A. *Crystallization and data collection*				
Initial conc. of duplex DNA (mM)	0.24	0.24	0.17	0.17
Initial conc. of divalent cation (mM)	8.57	8.57	61.20	31.40
Initial conc. of streptonigrin (mM)	0.11	0.11	0.00	0.28
Initial conc. of spermine-HCl (mM)	0.24	0.00	0.00	0.00
Final conc. of MPD of reservoir (%)	45	45	30	30
No. of cations/DNA phosphate	2.0	2.0	20.0	10.5
Exposure time/image (sec)	30.0	30.0	60.0	60.0
Oscillation angle step(°)	2.0	2.0	2.0	2.5
Total oscillation angle(°)	192.0	180.0	170.0	190.0
No. of reflections[a]	75660(13619)	77354(13464)	62360(13423)	63745(13087)
Completeness, all data (%)[b]	97.5(94.9)	95.0(90.7)	95.5(91.5)	93.3(89.5)
I/σ[b]	25.0(9.1)	21.6(7.4)	30.7(9.1)	32.8(6.9)
R_{merge}, all data (%)[b]	4.4(10.4)	4.7(12.3)	3.2(9.0)	3.1(12.2)
Mosaicity(°)	0.64	0.90	0.45	0.55
Unit cell parameters, space group monoclinic C2				
a(Å)	31.85	32.00	31.75	32.24
b(Å)	25.10	25.37	25.37	25.35
c(Å)	34.14	33.63	34.09	34.19
β(°)	114.87	112.98	116.88	117.17
B. *Structure refinement statistics*				
R_{factor}, all data (%)[b]	11.83(12.90)	14.09(15.29)	12.14(13.90)	14.05(16.53)
R_{free}, all data (%)[b]	14.05(19.66)	15.92(23.95)	14.77(16.29)	15.09(18.40)
Rms bond distance(Å)[c]	0.019(0.022)	0.033(0.018)	0.018(0.019)	0.018(0.018)
Rms bond angle (Å)[c]	2.91 Å(0.046)	2.98 Å(0.049)	2.73 Å(0.048)	3.61 Å(0.053)
Total no.of heavy atoms[d]	374(270.8)	327(272.8)	352(276.5)	343(271.4)
No. of solvent atoms[d]	109(68.8)	93(70.8)	126(74.5)	99(69.4)
No. of cation complexes[d]	4(2.3)	4(3.2)	4(3.0)	6(3.9)
No. of restraints	5407	3977	4814	5024
No. of parameters	3383	2945	3129	3052
Data/parameter	5.6	5.9	5.8	5.9
B-value[e]				
DNA residues	6.35(2.01)	9.09(2.88)	6.64(2.54)	8.55(3.42)
Solvent residues	12.45(6.12)	17.03(5.76)	11.93(5.29)	15.11(5.37)
All residues	8.14(4.62)	11.35(5.30)	8.53(4.53)	10.45(5.04)
Anisotropy[e]				
DNA residues	0.337(0.134)	0.441(0.141)	0.418(0.146)	0.409(0.124)
Solvent residues	0.306(0.171)	0.336(0.131)	0.373(0.141)	0.406(0.160)
All residues	0.328(0.146)	0.411(0.146)	0.402(0.145)	0.408(0.135)

[a] Total number collected and (number of unique reflections after processing).
[b] Values for all data and (data in last resolution shell, 1.016-0.985 Å).
[c] Rms deviation from parameters of Parkinson *et al.* (1996), calculated in XPLOR. The corresponding deviations from shelxl (DFIX for bond length and DANG for bond angles) are given in parentheses. Note that the unit for bond angle deviation in XPLOR is degrees. XPLOR bond length deviations are larger for CGMg because bond length and bond angle restraints for base atoms were not used in the last stages of shelxl refinement in CGMg. The rms deviations for restrained atoms (shelxl values), however, are comparable to those of the other three structures.
[d] Each atom record counts as one atom, so a disordered atom with two conformations is counted here as two atoms. The number in parentheses is the total atom occupancy.
[e] Heavy atoms only. Value in parentheses is standard deviation.

pairs 3 and 4 in AGMg. At the helix junctions, where base sequence is conserved and the minor groove width varies by no more than 1 Å between structures, only one conformation is present regardless of cation concentration: Mg111 sits directly on the 2-fold axis in CGMg and AGMg, while Ca111 sits ca. 1 Å off the 2-fold axis in CGCa, AGCa and CICa.

Minor groove width variation cannot of itself predict DNA bending

Figure 4 compares normal vector plots for the Mg^{2+} and Ca^{2+} forms of sequences *CG*, *AG* and *CI*. These can be regarded as maps of change of orientation of vectors perpendicular to the base-pairs, viewed down the helix axis. Figure 5 plots minor and major groove widths as calculated by CURVES (Lavery & Sklenar, 1988) and plotted with SHELXDNA. Figure 6 compares helix-independent roll, twist, propeller twist and tilt as calculated by FREEHELIX (Dickerson, 1998; Dickerson & Chiu, 1997). The normal vector plots reveal a common pattern of a smooth, planar bend across the junction between stacked helices, at base-pairs G9, G10, C1 and C2. Another view of this strong inter-helix bending is provided by the three successive large positive roll angles (compression of the major groove) along the sequence GGCC at left and right of Figure 6(a). The compensating reverse bend at

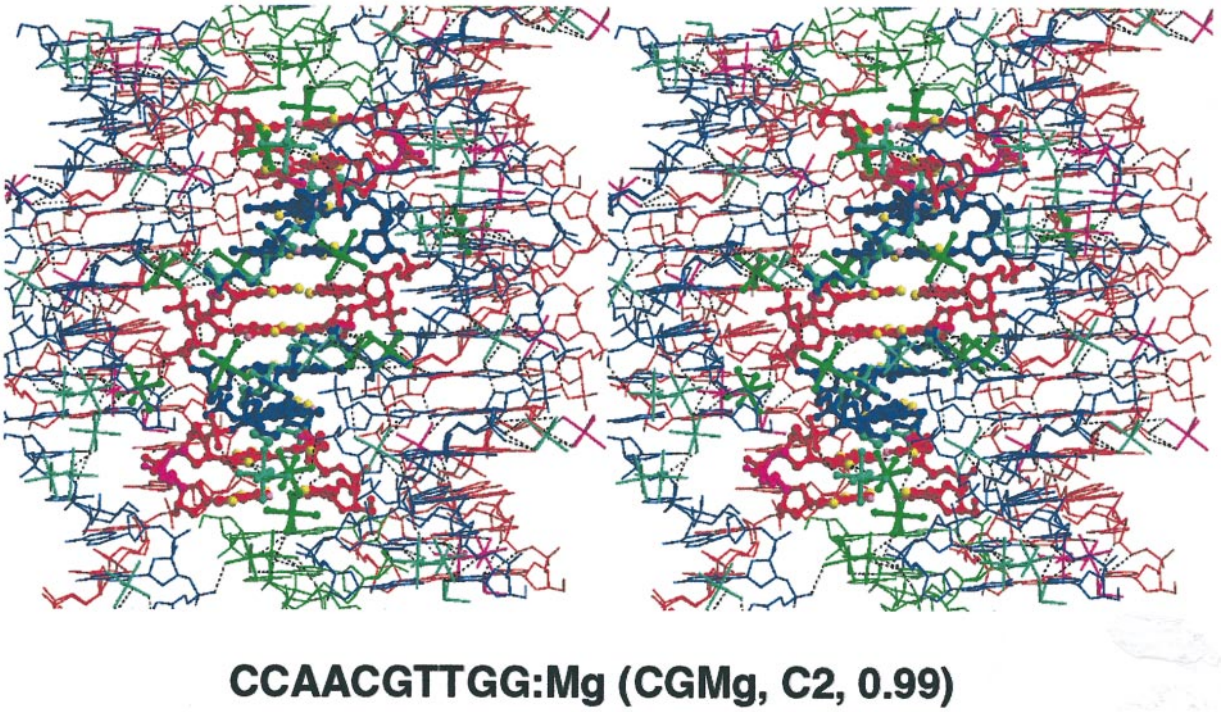

Figure 2. Stereograms for CGMg (a) and CGCa (b) as in Figure 1, now with the addition of neighboring molecules in thin line. Broken lines denote H-bonds or Van der Waals contacts. Prepared using Molscript/Raster3d/ImageMagick. Crystal packing for AGMg and AGCa are similar and not shown.

the center of the decamers is less regular, but is embodied in an alternation of roll values at successive steps. As is always found, the contribution of tilt to helix bending is negligible (Figure 6(d)). This tendency for roll-bending at the sequence TGGCCA was first noted by Goodsell *et al*. (1993). They observed it at the center of their CATGGCCATG helix and then observed that the bend was universal across the inter-helix junctions in other stacked-helix sequences of the general form: CCAxxxxTGG. These junctions, consisting of the last three base-pairs of one helix and the first three of the next, are compared in Figure 7, where the bending at the innermost four of these six base-pairs is obvious. As was mentioned earlier, this bending occurs around cations 113 within the major groove at the helix terminus.

The CIMg sequence is an anomaly compared to the foregoing behavior. Its normal vector plot displays the ellipse that is the mark of smooth, continuous writhe, without bending excursions (Figure 4). It also does not have the large roll excursions, positive or negative, that are found in the other decamers (Figure 6(a)). The reasons for this anomalous behavor may lie in its unique crystal packing. In constrast to the parallel side-by-side packing of helix columns in the monoclinic C2 decamers (Figure 2), columns of helices in the CIMg decamer are packed at an acute angle relative to one another. At bases 8-10, the major groove nests against the phosphate backbone of a neighboring helix. (See Figures 5 and 6 of Lipanov *et al*., 1993.) As a consequence, the cation-induced bend observed at bases 9-10 in the C2 decamers is completely obliterated. At the other end of the helix, the absence of bending at bases 1 and 2, where there is no groove-against-backbone packing, is consistent with an absence of bound Mg^{2+} at bases 19 and 20, although it also is possible that the Mg^{2+} was not observed because of the low resolution of X-ray data.

Alternative backbone conformations and thermal ellipsoids of individual atoms are shown in Figure 7. (See also Figures 1 and 2 for alternative conformations.) Most of the conformational heterogeneity in DNA resides in the sugar-phosphate backbone. At the central two base-pairs, the backbone flexibility gives rise to large variations in minor groove width but the orientation of base-pairs remain constant. Similarly, large variations in minor groove width occur at base-pairs 2-3 and 8-9 even though there is very little variation in roll and twist angles. Lastly, a large variation in roll angle ($\sim 8^\circ$) between conformers at the central CG step of CGCa cannot be attributed to the width of the minor groove, which is nearly identical in the two conformers. In this case, the backbone is trapped by four ionic bonds between Ca114 and O1P and O2P atoms of bases 6-7 and 16-17 of neighboring helices. Hence, the inherent flexibility of the DNA backbone makes the minor groove width an unreliable parameter for indicating DNA bending.

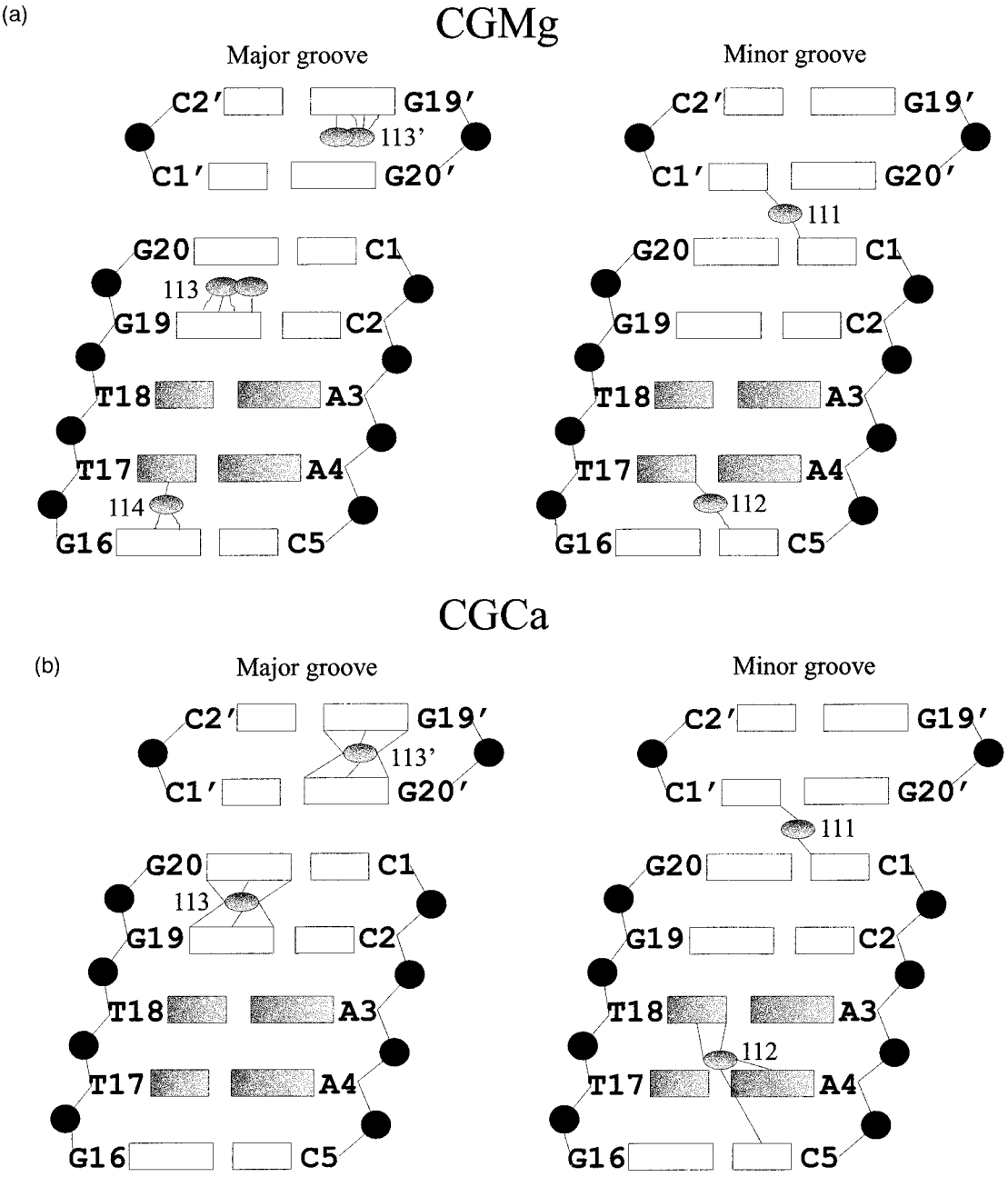

Figure 3 (*legend shown on page 928*)

Figures 1(b) and 4-5 also indicate that the two conformations for bases 5 and 15 in CGCa correspond to the most favored conformations in this sequence. One conformation is observed in CGMg and *AG* decamers, while the other is observed in CIMg.

Minor groove width affects its hydration property

A direct consequence of the backbone flexibility is a dependence of minor groove hydration on its width. A narrow minor groove gives rise to a highly ordered spine of hydration which is two layers deep (see Figure 2 of Chiu *et al*., 1999): a water molecule in the first layer sits between adjacent base-pairs and makes four H-bond interactions, two to base atoms and two to water molecules in the second layer. As the groove narrows, a third hydration layer, with its water molecules contacting O4′ atoms of the groove walls, is observed (Shui *et al*., 1998b; Tereshko *et al*., 1999ab). Similarly, as the groove widens, as from CGMg to CGCa, two well-ordered ribbons of hydration which run side-by-side appear, with the

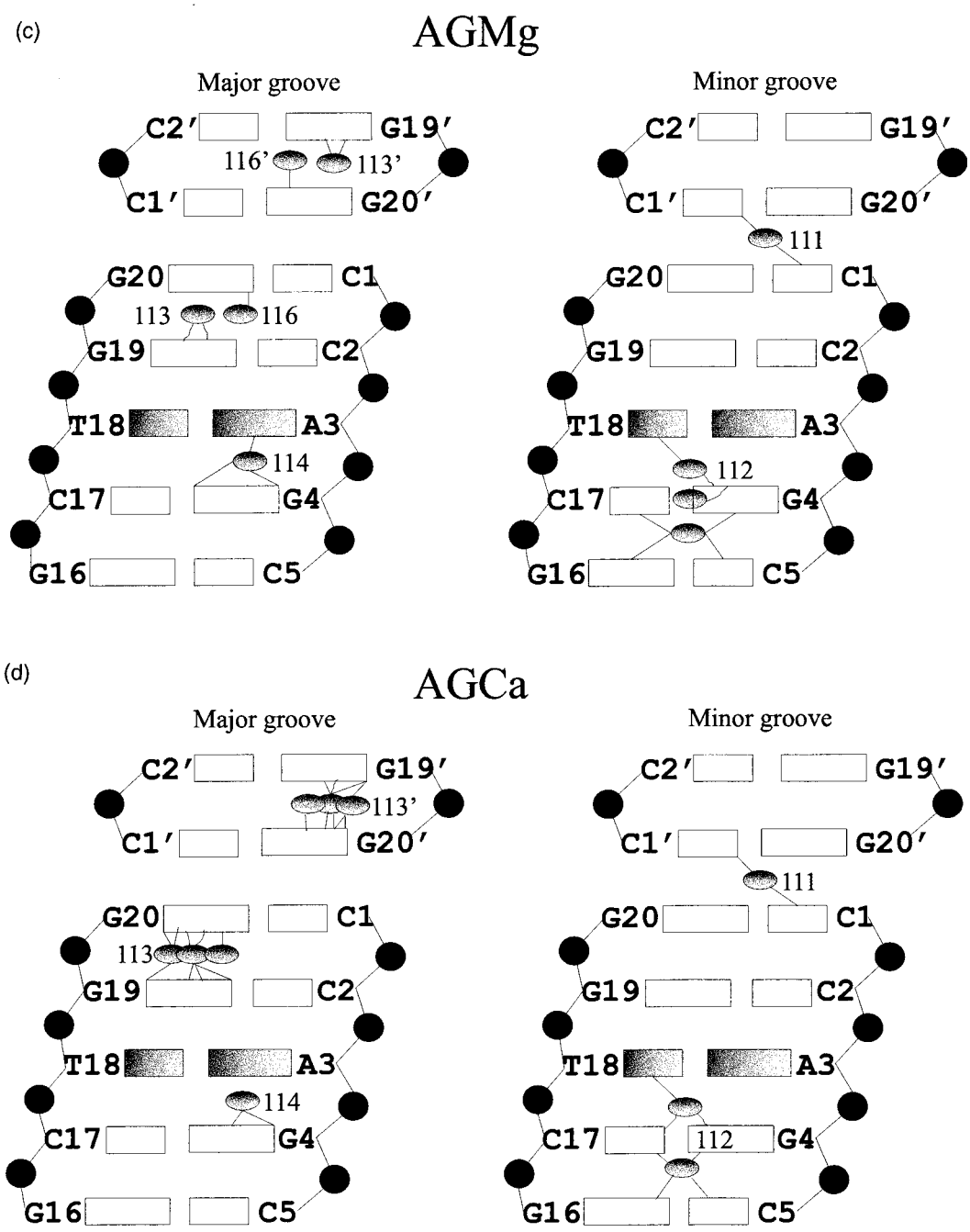

Figure 3 (*legend shown on page 928*)

water molecules in the first hydration layer sitting directly atop each base atom (green ribbons in Figure 1(a)-(b)). These spines are essentially continued by the bound cation complex 112 towards the ends of the helix, where the wide groove is again hydrated by two ribbons. Likewise, the narrow minor groove of the central two base-pairs in AGCa is occupied by one water molecule whereas the wider groove at the same base-pairs in AGMg has no bound water (Figure 1(c)-(d)).

The hydration patterns of these four structures, as well as the structural information derived from the conformational disorder and anisotropy of backbone atoms (see previous section) provide a simple model for the passive binding of bulky hydrated divalent cations and site-specific drugs to the minor groove of *B*-DNA: the narrow minor groove can accommodate the binding of drugs such as lexitropsins in monomeric form, but the flexible backbone atoms can easily swing open to

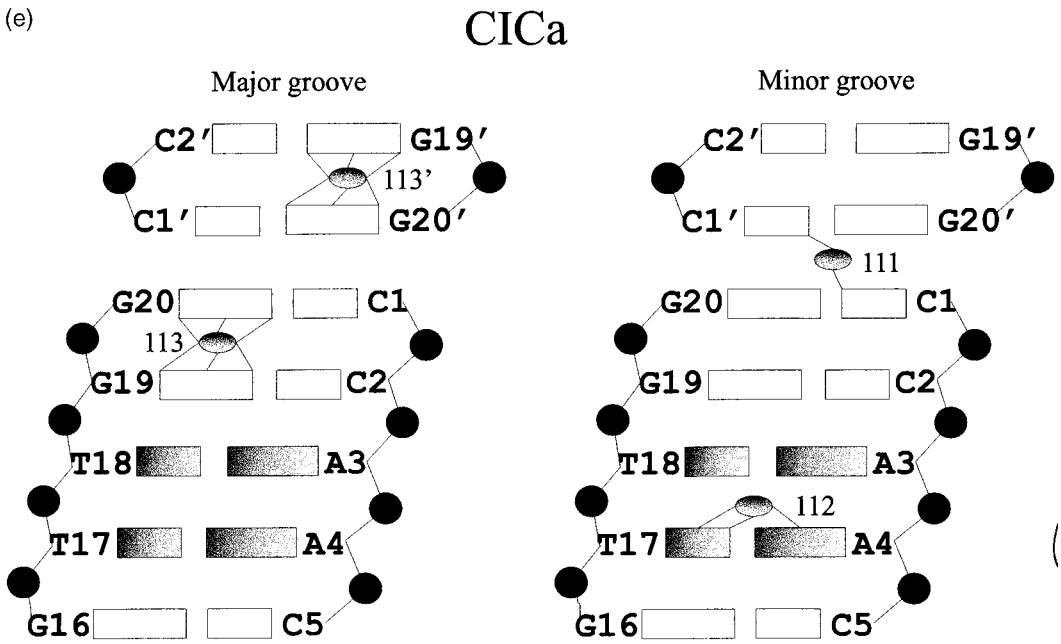

Figure 3. Schematic diagram of cation interactions to DNA base atoms. Only interactions to base atoms are shown. The cations are shown as grey ellipses and each interaction is shown as a single thin line between the ellipse and base. Only one half of the duplex, and two base-pairs from a symmetry-stacked duplex, are shown. Refer to Figures 12 and Table 3 for specific interactions.

accommodate their binding in side-by-side modes as well (see Kopka *et al*., 1998, and references therein). In contrast, the binding of the much larger protein ''ligands,'' such as TATA-Binding-Protein, to the minor groove is a non-passive process and leads to the drastic local unwinding of DNA (see Juo *et al*., 1996, and references therein).

Major groove binding bends DNA by base roll compression towards the major groove

As Figures 6(a) and 8 demonstrate, the binding of cations to the major groove at GG bases 9-10 and 19′-20′ across the helix junction arises from three consecutive positive roll angles which compress the major groove and produce a smooth planar curvature of the order of 10-15 Å for Mg^{2+} and $\sim 20^\circ$ for Ca^{2+}. Curvature is greater for Ca^{2+} than for Mg^{2+}, probably because of the geometry of calcium binding to DNA base atoms: Ca^{2+} locks the GG bases by polar covalent bonds whereas Mg^{2+} interacts through its water ligands. The polar covalent bonds induce a greater negative propeller twisting of the terminal base-pairs (Figure 6(c)), which aids in base stacking. This cation-induced bend is neither an artifact of crystal packing nor of the missing covalent linkage between stacked helices, as it is also present in CCAAGATTGG:Mg^{2+} (bdj008), CCAAGCTTGG:Ca^{2+} (bdj052), and CATGGCCATG (bdj051; Goodsell *et al*., 1993), all of which are non-isomorphous with the C2 decamers (Figure 8). Of the groove cations, only Mg112 and Mg113 were observed in bdj008, and Ca111 in bdj052. The other cations at the conserved base-pairs are probably missing because of the lower resolution and higher data collection temperature.

Thermal stabilization of DNA structure by bound divalent cations

Two statistics relevant to the discussion of cation-dependent stabilization of DNA structure are *B*-value and anisotropy. *B*-value is a measure of the displacement of an atom about its average position due to thermal vibration and static disorder. For reference, values of 100, 10 and 1 $Å^2$ correspond to mean atomic displacements of 1.60, 0.36 and 0.11 Å, respectively. Anisotropy is defined as ratio of the minimum and maximum eigenvalues of the 3×3 matrix of ADPs. A value of 1 means that an atom moves in all directions with equal probability, and a value less than one means that motion in one direction is favored over all others. Table 4 gives the *B*-values and bond distances for the Mg^{2+} and Ca^{2+} complexes. Note the uniformity of *B*-value and bond length characteristic of well-ordered bound cations.

Under the same crystalline environment, the lower overall *B*-values of the Ca^{2+} structures indicate that Ca^{2+} binding stabilizes DNA more than does Mg^{2+} binding (Figure 9): each DNA atom of the Ca^{2+} and Mg^{2+} structures has a mean displacement of 0.28 Å and 0.34 Å about

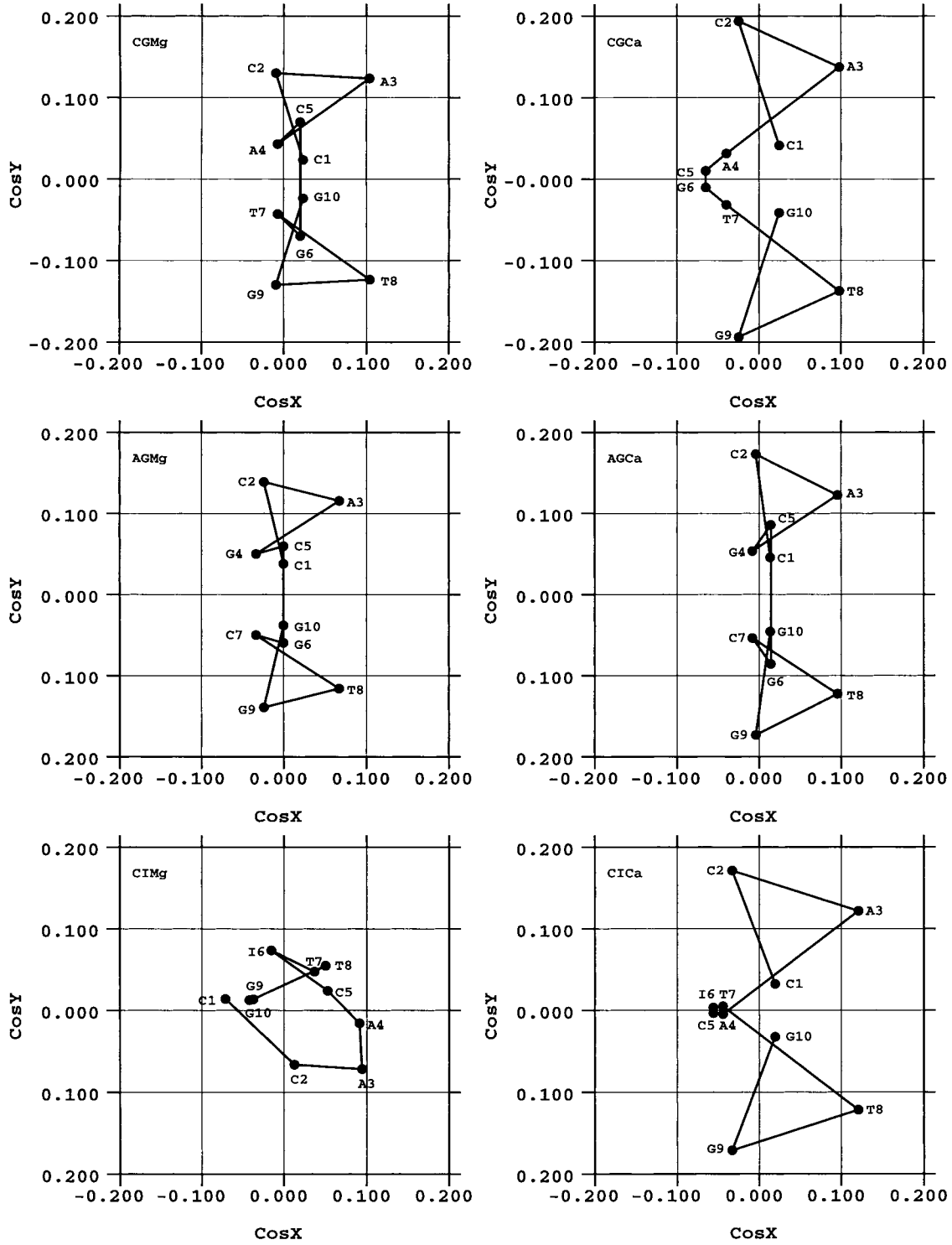

Figure 4. Normal vector plots for (a) CGMg, (b) CGCa, (c) AGMg, (d) AGCa, (e) CIMg and (f) CICa. Labeled points mark the tips of unit vectors normal to the base-pairs, viewed down the helix axis. A linear sweep of points across the plot indicates a bend in helix axis, and a circle or ellipse indicates writhe. For more details see Dickerson & Chiu (1997) and Dickerson (1998).

its equilibrium average, respectively. This stabilization is local, with minimal motion around base-pairs 2-3 and 8-9. Comparing the average *B*-value of residue 2 between CGMg and CGCa, for example, tells us that residue 2 of CGMg is ~20% more disordered than that of CGCa. Similarly, in CGMg residue 5 is ~25% more disordered than residue 2.

Table 3. Sequence-specific cation-DNA interactions

CGMg					
111	OW5-O2//C1 (2.68)	OW5-O4′//C2 (3.11)	OW5′-O2//C1′ (2.68)	OW5′-O4′//C2′ (3.11)	
112	OW1-O2//C5 (2.87)	OW3-O2//T17 (2.47)	OW3-O4′//T18 (2.67)	OW6-O4′//G6 (2.65)	
113a	OW1-N7//G19 (2.80)	OW2-O1P//C5′ (2.96)	OW5-O6//G19 (2.85)		
113b	OW5-N7//G19 (2.71)	OW1-O2P//G19′ (2.76)			
114	OW2-N7//G16 (2.84)	OW3-O4//T17 (2.61)	OW4-O2P//T7b′ (2.79)	OW4-O2P//T7′ (2.87)	OW5-O6//G16 (2.81)
	OW6-O1P//T7b′ (3.14)	OW6-O2P//T7′ (3.06)			
CGCa					
111	OW1-O2P//G20′ (2.97)	OW2-O1P//C2b′ (2.87)	OW2-O1P//G20′ (2.91)	OW3-O2//C1′ (2.70)	OW4-O3′//C1b′ (2.99)
	OW4-O4′//C1b′ (3.01)	OW6-O2//C1 (2.92)	OW6-O4′//C2 (2.70)		
112	OW1-O4′//T18 (2.87)	OW2-N3//A4 (2.76)	OW4-O2//T18 (2.77)	OW4-O4′//G19 (2.88)	OW5-O2//C5b (3.06)
	OW5-O4′//C5 (2.82)				
113	**CA-N7//G19 (2.67)**	**CA-O6//G20 (2.55)**	OW2-N7//G19 (3.15)	OW2-O2P//G19 (2.87)	OW3-N7//G20 (2.67)
	OW4-O6//G19 (2.61)	OW5-O6//G20′ (2.93)	OW5-O6//G20 (2.62)		
114	***CA-O1P//T17a′ (2.35)***		***CA-O2P//G6b (2.31)***	OW3-N7//G6 (3.08)	
AGMg					
111	OW1-O2//C1 (2.69)	OW1-O4′//C2 (3.07)	OW1′-O2//C1′ (2.69)	OW1′-O4′//C2′ (3.07)	
112a	OW6-O2//C5 (3.01)	OW6-N2//G4 (3.15)	OW1-O2//C17 (2.56)	OW1-O4′//T18 (2.82)	OW3-N7//G16 (2.92)
	OW3-O4′//G6 (2.86)				
112b	OW1-N2//G4 (3.24)				
112c	OW3-N2//G4 (3.24)	OW1-O2//T18 (2.99)	OW1-O4′//G19 (2.64)		
113	OW5-N7//G19 (2.77)	OW6-O1P//C5′ (2.47)	OW3-O6//G19 (2.75)		
116	OW1-N7//G20 (2.72)				
114	OW1-N7//A3 (2.78)	OW1-O2P//A3b (3.11)	OW1-O2P//A3 (2.55)	OW2-O1P//T7b′ (2.53)	OW3-N7//G4 (2.73)
	OW4-O1P//T7b′ (3.06)	OW4-O2P//T7c′ (3.12)	OW4-O3′//G6b′ (2.46)	OW4-O2P//T7a′ (2.66)	OW5-O6//G4 (2.65)
	OW6-O1P//G6′ (2.64)				
115	OW5-O1P//G6 (2.62)	OW5-O2P//A3a′ (2.75)	OW5-O2P//A3b′ (2.85)		
AGCa					
111	OW5-O1P//C2′ (2.86)	OW6-O2//C1′ (2.60)	OW4-O3′//C1′ (2.91)	OW2-O2//C1 (2.97)	OW2-O4′//C2 (2.68)
	OW3-O2P//G10′ (2.89)				
112a	OW1-O4′//T18 (2.89)	OW1-O2//C17 (3.00)	OW6-N2//G4 (3.21)	OW6-O2//C5 (2.78)	OW4-O2//T18 (2.89)
	OW4-O4′//G19 (2.98)	OW2-O2P//C2a′ (2.91)			
112b	OW1-O4′//T18 (2.81)	OW1-O2//C17 (2.64)	OW4-N2//G16 (3.21)	OW4-O4′//G6 (2.77)	OW6-O2//C5 (2.65)
	OW7-N2//G4 (3.21)				
113a	**CA-N7//G19 (2.71)**	**CA-O6//G20 (2.64)**	OW5-N7//G19 (3.10)	OW5-O2P//G19 (2.85)	OW4-N7//G20 (2.77)
	OW2-O6//G20 (2.61)	OW2-N4//C1 (3.21)	OW1-N4//C1 (3.26)	OW1-O6//G19 (2.55)	
113b	OW3-N7//G20 (2.77)	OW2-O2P//G19 (2.85)	OW1-O1P//G4′ (2.74)		
113c	OW1-O6//G20 (3.01)	OW1-N7//G20 (2.52)	OW4-O6//G19 (2.83)	OW6-O1P//C5′ (2.69)	
114	OW2-O2P//A3 (2.67)	OW3-N7//G4 (2.68)	OW1-O6//G4 (2.77)	OW2-O1P//G6b′ (2.56)	OW2-O1P//G6′ (2.95)

Roman, not bold, hydrogen bonds. Roman, bold face, polar-covalent bonds. Italic, bold face, ionic bonds. OW2-O6//G19 (2.61), Bond from water oxygen 2 to the O6 of base G19, bond length 2.61 Å. CA-N7//G19 (2.71), Bond directly from calcium atom to the N7 atom of base G19, bond length 2.71 Å. Values in parentheses are bond distances in Å. Only one asymmetric unit is tabulated, corresponding to the top half of Figure 1. The bottom half is related by a 2-fold axis. Atoms in symmetrically related neighboring molecules are marked by a prime (′).

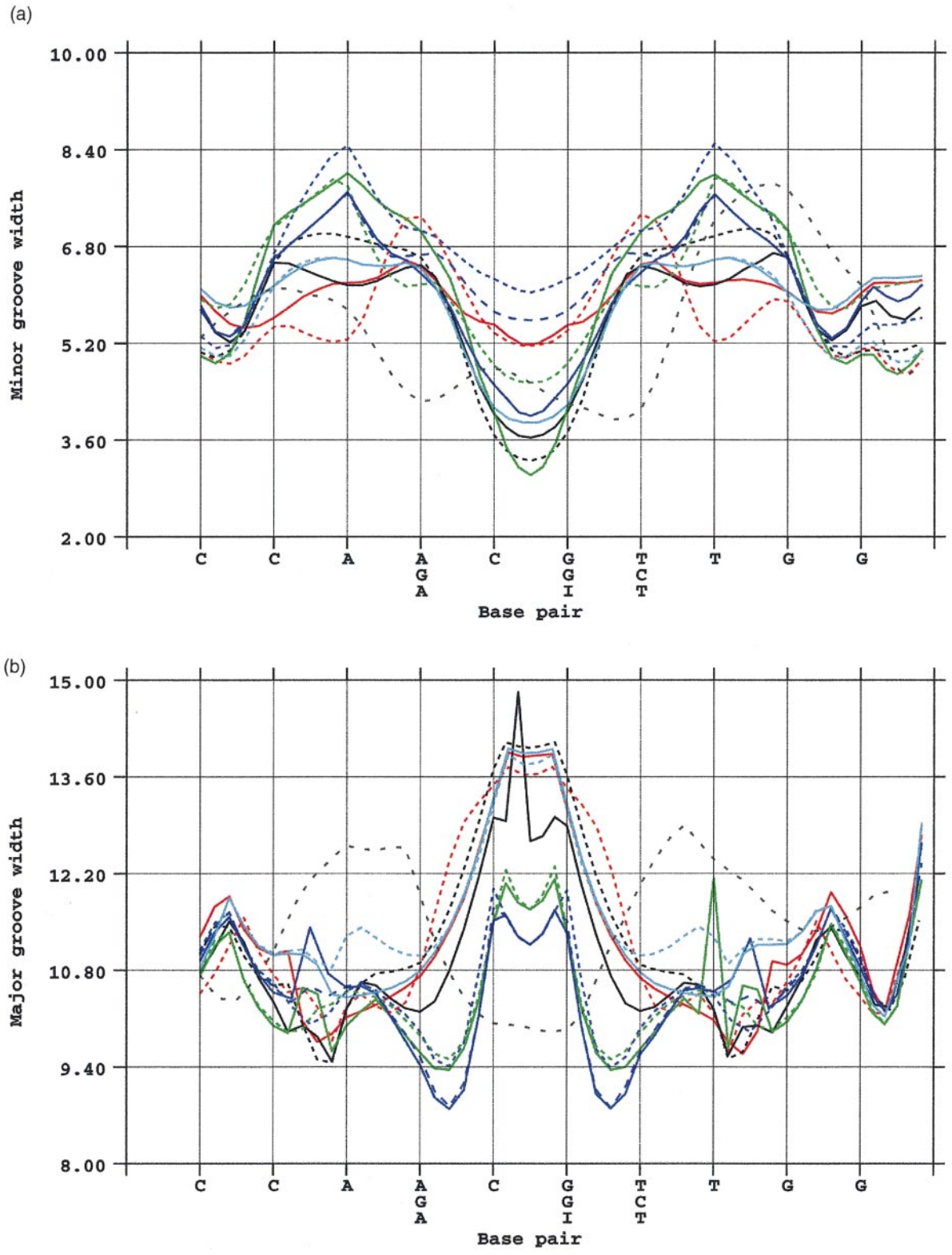

Figure 5. Groove widths as calculated by CURVES and plotted by SHELXDNA: (a) minor groove, (b) major groove. Color codes are: black, CGMg; red, CGCa; blue, AGMg; green, AGCa; light blue, CICa; and grey, CIMg. Base sequences are labeled along the x-axis from top to bottom in this same order. Broken lines denote alternative conformations.

Table 4. *B*-values and bond lengths in cation complexes

Ion	Occu.	Metal	OW1	OW2	OW3	OW4	OW5	OW6	OW7	C1[e]	C2[e]	⟨*B*-value⟩	⟨*R*⟩
CGMg												**15.18**	**2.08**
111	1.00*	10.15	13.23		13.57		13.94					12.72	
			2.05		2.12		2.09						2.09
112	0.66	11.68	14.05	12.91	10.46	17.01	16.29	10.35				13.25	
			2.07	2.08	2.07	2.07	2.09	2.07					2.08
113a	0.61	11.24	18.42	13.44	14.57	11.63	8.35	16.57				13.46	
			2.08	2.07	2.06	2.07	2.13	2.09					2.08
113b	0.39	9.33	13.12	14.69	12.14	16.71	11.07	12.82				12.84	
			2.08	2.09	2.09	2.08	2.08	2.08					2.08
114	1.00	18.37	24.78	18.67	25.90	25.02	16.05	29.31				22.58	
			2.07	2.05	2.06	2.09	2.09	2.11					2.08
CGCa												**7.33**	**2.43**
111	0.50*	4.18	4.42	5.09	5.84	4.86[e]	6.76	4.72	5.08	4.72[e]	3.04[e]	4.87	
			2.46	2.50	2.42	2.44	2.37	2.39	2.42				2.43
112	0.64	8.98	8.49	9.81	13.66	7.88	11.90	11.23	10.74			10.34	
			2.46	2.41	2.41	2.38	2.39	2.39	2.50				2.42
113	0.62	4.81	7.36	7.19	6.01	5.63	6.59	4.46[a]	5.21[b]			5.91	
			2.41	2.44	2.47	2.38	2.38	2.61	2.55				2.46
114	0.50	7.72	8.42	9.33	8.13	10.13	6.23[c]	9.26[d]				8.46	
			2.39	2.38	2.42	2.45	2.35	2.31					2.38
AGMg												**12.50**	**2.08**
111	1.00*	6.77	10.04		9.46		9.66					8.98	
			2.13		2.09		2.06						2.09
112a	0.36	10.74	6.17	10.88	18.37	11.31	12.90					11.73	
			2.10	2.08	2.08	2.09	2.07						2.08
112b	0.37	15.53	12.91	13.63	11.65	17.17	16.71					14.60	
			2.07	2.09	2.10	2.09	2.06						2.08
			2.09	2.09	2.07	2.10	2.07						2.08
113	0.76	10.39	9.96	14.09	8.23	12.91	12.34	14.03				11.71	
			2.08	2.09	2.08	2.09	2.11	2.08					2.09
114	1.00	16.99	14.11	16.90	16.54	16.91	21.85	21.20				17.78	
			2.08	2.06	2.09	2.06	2.08	2.08					2.08

115	0.21	7.09	6.21	9.16	6.64	10.79	5.42	8.02				7.62	
			2.08	2.09	2.08	2.08	2.09	2.08					2.08
116	0.414	14.47	13.48		10.64		11.48	16.89				13.39	
			2.07		2.05		2.08	2.09					2.07
AGCa												**8.42**	**2.45**
111	0.50*	5.01	5.93	5.27	4.85	5.14 [e]	6.28	5.77	5.42	6.13[e]	4.22[e]	5.40	
			2.41	2.38	2.45	2.45	2.43	2.48	2.41				2.43
112a	0.58	6.13	7.87	7.49	8.50	6.57	10.23	8.81	7.11			7.84	
			2.46	2.42	2.44	2.46	2.44	2.44	2.43				2.44
112b	0.42	9.65	8.81	11.17	10.81	10.60	11.84	10.73				10.52	
			2.46	2.42	2.44	2.44	2.42	2.41	2.44				2.43
113a	0.38	5.46	7.10	7.16	8.00	6.67	7.50	4.96[a]	5.53[b]			6.55	
			2.46	2.41	2.44	2.46	2.43	2.71	2.64				2.51
113b	0.31	9.10	8.91	9.13	11.54	10.26	13.27	9.10				10.19	
			2.45	2.46	2.48	2.46	2.47	2.47					2.47
113c	0.31	7.41	9.54	8.66	9.57	5.57	8.12	9.60				8.35	
			2.44	2.44	2.45	2.44	2.45	2.44					2.44
114	0.50	14.24	8.98	14.53	11.45							12.30	
			2.41	2.44	2.42								2.42

Ion, Hydrated ion as numbered in the indicated DNA crystal structure.
Occu., Crystallographic occupancy of the central metal ion site.
*Note: the metal atom sits on a twofold symmetry axis in CGMg and AGMg, but not in CGCa and AGCa.
Metal, *B*-value for the central metal cation.
OW1-OW7, *B*-values for coordinated O or N atoms, with distance from central metal cation listed below in Å.
Columns ⟨*B*-value⟩ and ⟨*r*⟩ at right are mean values of *B* and distance from ligand to metal atom.
Special atom labels: [a]N7 atom on base A/G9; [b]O6 atom on base G10; [c]O2 atom of phosphate 6; [d]O1 atom of phosphate 17 on neighboring molecule; [e]O of liganded ethanol molecule.

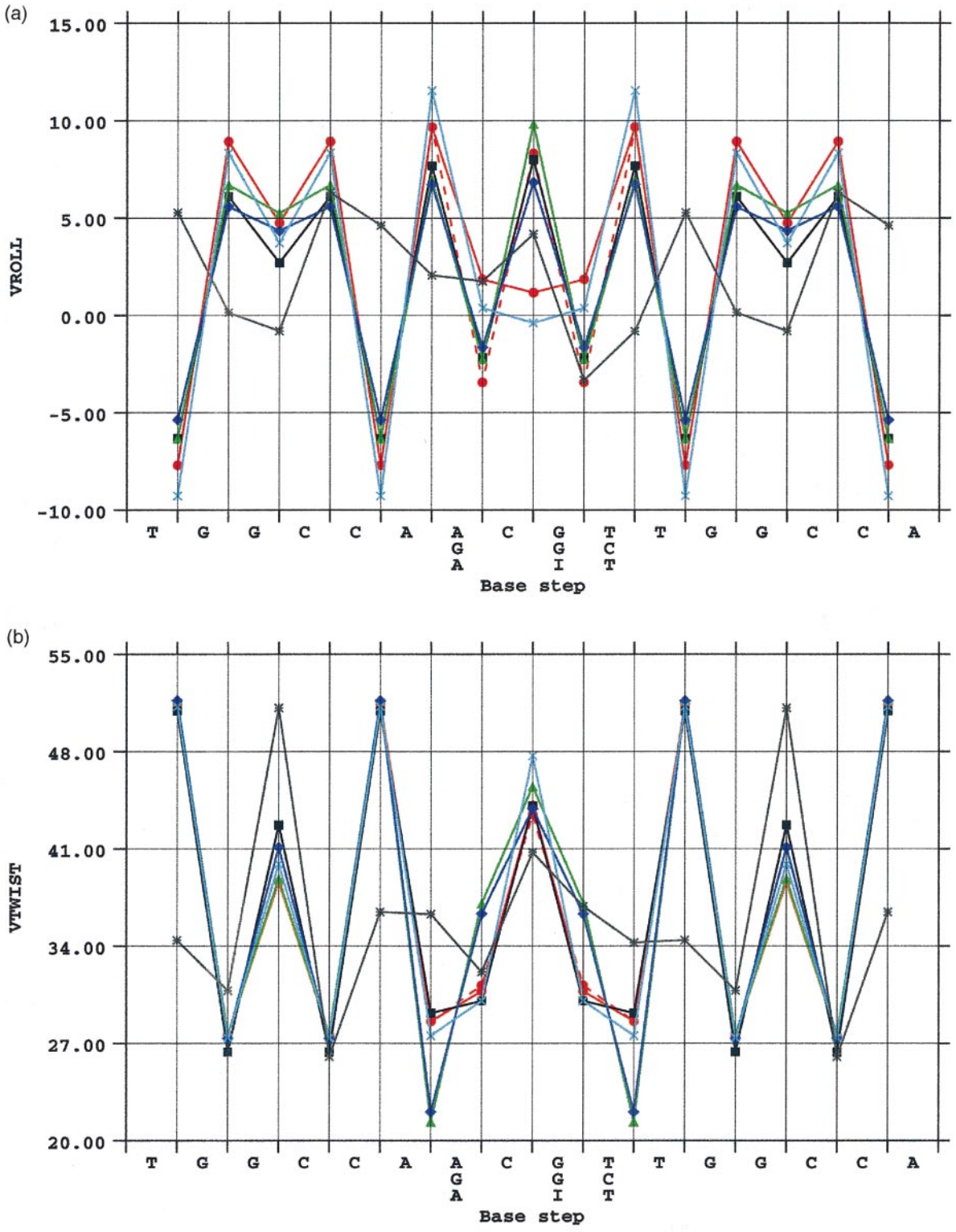

Figure 6 (*legend opposite*)

Figure 7 shows the anomalous motion of each atom in CGCa. The Figures for CGMg, AGMg, and AGCa are similar and not shown. Note that the most flexible part of the DNA structure, the sugar-phosphate backbone, is the most anisotropic and disordered as expected. Alternate conformations are clearly observed for the sugar-phosphate backbone, with the direction of the thermal ellipsoids being perpendicular to the helix axis, pointing in the direction of opening and closing of the minor groove. This motion leads to large variations in minor groove width

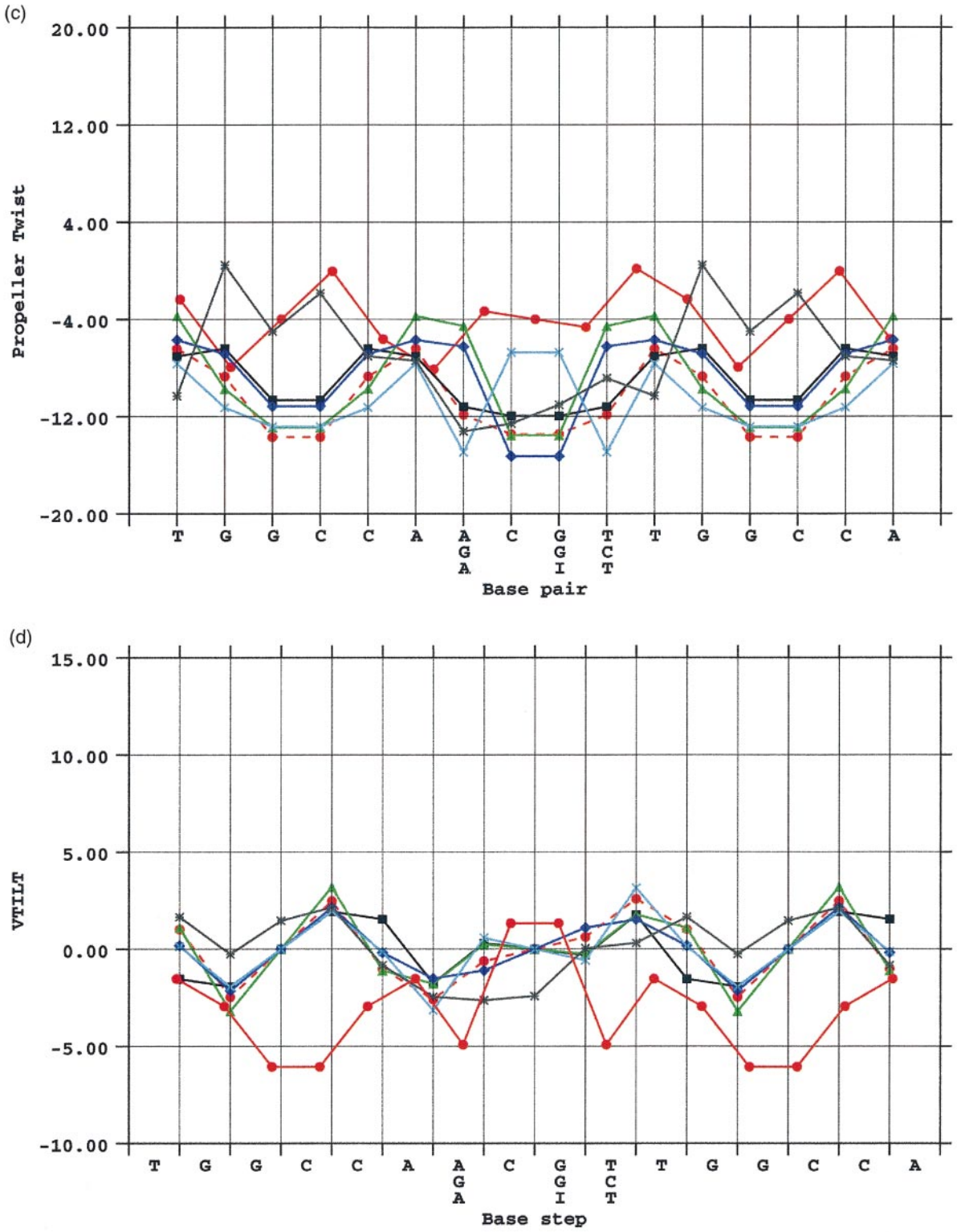

Figure 6. Helix parameters as calculated by FREEHELIX and plotted by SHELXDNA. (a) Roll, (b) twist, (c) propeller and (d) tilt. Helices are identified as in Figure 5.

which do not necessarily imply DNA bending (see previous section also). In contrast, atoms in individual base-pairs have thermal ellipsoids that are elongated in the base-pair plain, particularly in the direction of the base-pair long-axis. Note that this anisotropy is less in base-pairs closer to the ends of the helix, because out-of-plane motion becomes easier there.

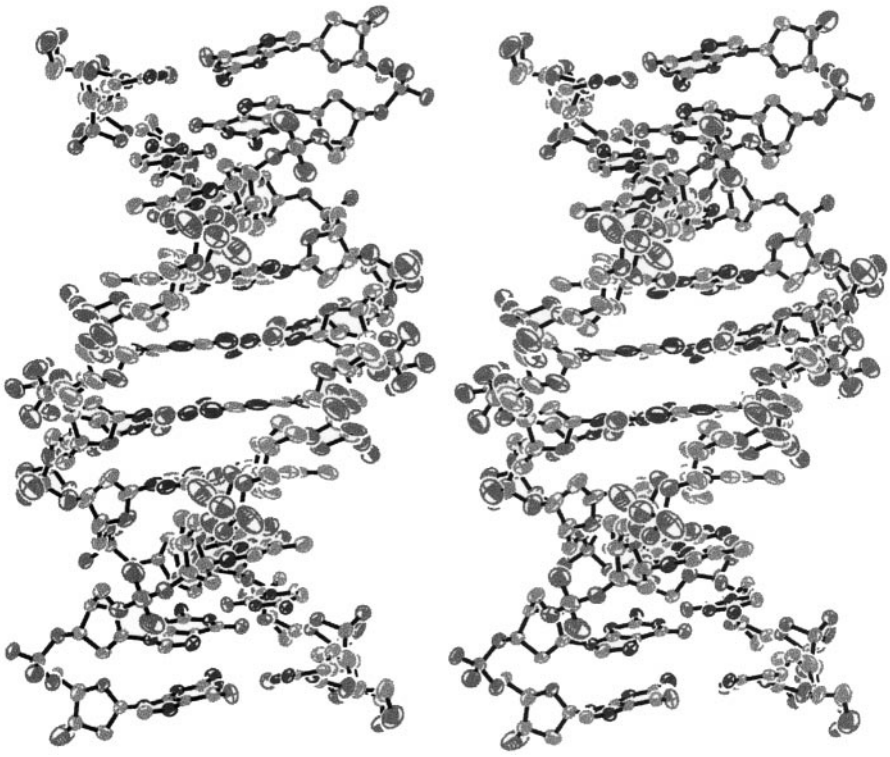

Figure 7. Ortep stereo plots of CGCa, generated by Shelxtl version 5.1 as distributed by Bruker AXS. Views into the minor groove at center. Note the anisotropy of backbone chains indicative of opening and closing the minor groove, and in-plane anisotropy of individual base-pairs. Two alternative configurations are used in some regions of backbone.

Discussion

Correlation between crystallization and data collection conditions, and number of observed cations

An issue of importance when studying DNA sequence-specific cation binding by X-ray crystallography is the resolution and quality of the data set. Does the absence of cation localization to a particular region in the crystal structure mean that it is not bound there? Not necessarily. An examination of a database of 28 cation-bound *B*-DNA structures spanning ten different crystal packing environments and three data collection protocols (Table 1 and Figure 10) shows that there is a correlation between experimental conditions and number of observed cations. Specifically, increasing buffer cation concentration in the crystallizing medium tends to force more localized cations in the crystal (Figure 10(a)) and more cations are observed at higher data resolution (Figure 10(b)). As mentioned earlier, divalent cations are the glue that holds the crystal lattice together. As the unit cell volume decreases, other factors being equal, the mean volume per atom or Matthews Number also decreases, and the number of lattice interactions rises. Consequently, the outer limit of resolution falls (''increased resolution'') and the number of observed cations per base-pair in the crystal structure rises as the Matthews Number falls (correlation values = 0.59 and 0.74, respectively). Hence the significance, number, and location of cations can be compared safely only between structures having similar crystallizing conditions, data collection methods and resolution.

Sequence-specific binding by Mg^{2+} and Ca^{2+}

Despite the correlation between number of observed cations and experimental conditions, the current available structures listed in Table 1 and plotted in Figure 10(c) show that cation binding to DNA is sequence-specific: the correlation coefficient for a linear fit of the cryogenic structures is 0.93. Note in Figure 10(c) that nine of the 16 cryogenic structures are overlapped in the seven points. This strong correlation between divalent cation binding and base sequence as observed in the 28 structures further supports the sequence-specific binding property observed in our four 0.99 Å structures: as discussed in Results, changing the base sequence of residue 4 directly leads to a change in cation binding, particularly in the major groove. The lower correlation coefficient of 0.74 for the fit between backbone binding and percent GC (Figure 10(d)) supports the non-specific nature of cation binding to the phosphate backbone, as discussed in Results.

Under identical crystal packing environments, and if the X-ray data are of sufficient quality (high

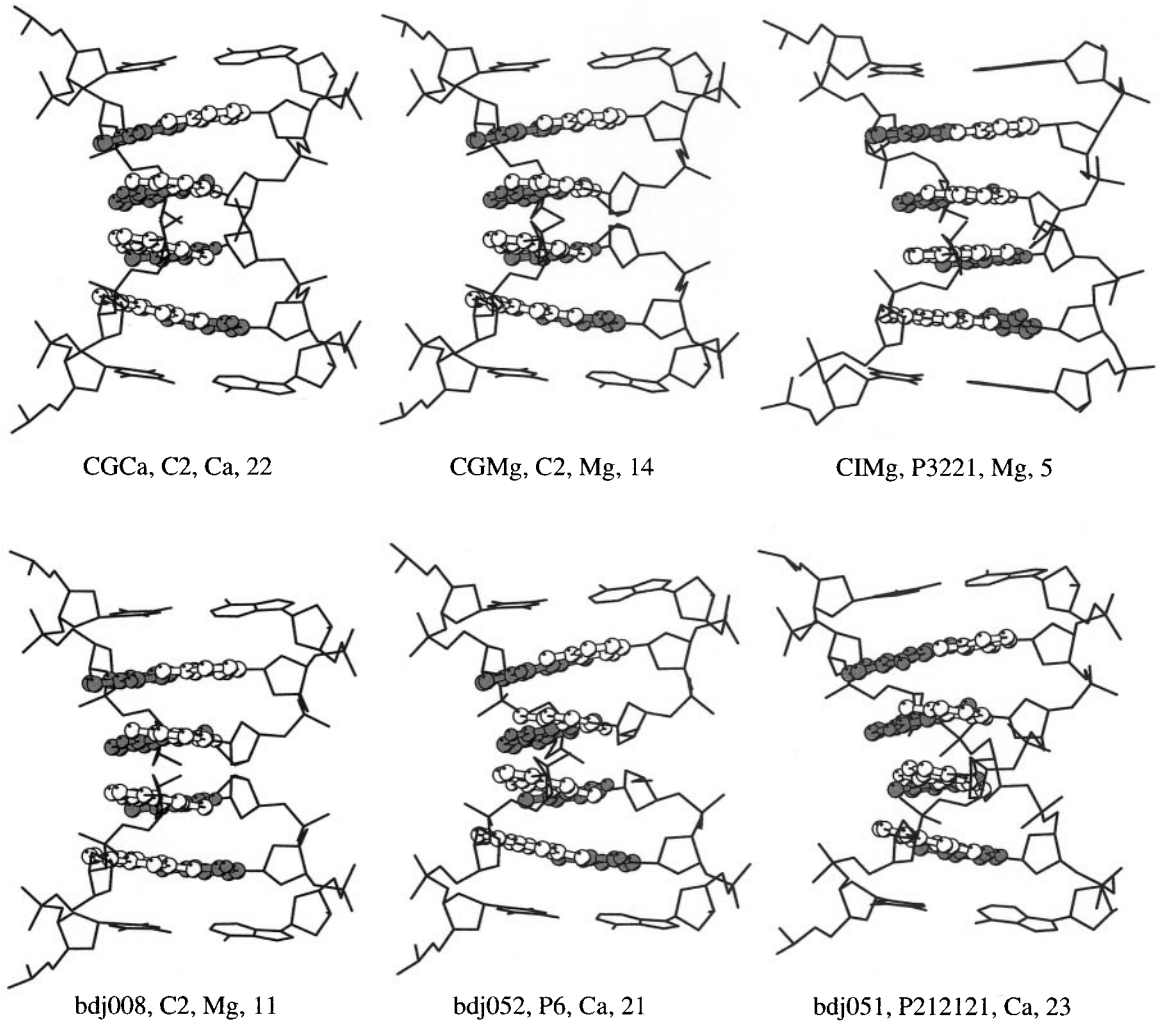

Figure 8. Closeup views of the TGG//CCA steps across helix junctions of CGCa, CGMg, CIMg, bdj008 (CCAAGATTGG) and bdj052 (CCAAGCTTGG), compared with the internal steps TGGCCA of bdj051 (CATGGCCATG). Labels are name, space group, cation, and angle in degrees between the second and fifth base-pairs as drawn here. Closeup views of AGCa, AGMg, and CICa are not shown but the bend angle between the second and fifth base-pairs are 19°, 15°, and 19°, respectively. With the exception of CIMg, all structures shown or mentioned here have a bound major groove cation at the GG steps. Taking into account the intrinsic bend angle in the cation-free CIMg structure, the binding of magnesium and calcium to GG steps results in an increased bend of ~10° and ~15° degrees, respectively. See text also.

resolution, completeness, and I/σ, and low R_{merge} and mosaicity) one might use the cation's occupancy or its order of peak assignment, to rank binding affinity. For the four 0.99 Å structures reported here, the minor groove binding affinity for Mg^{2+} is G-G > A-G > A-C and for Ca^{2+} is G-G > A-T ~ A-C, with cations positioned in the center of the groove. For the major groove, the order for Mg^{2+} is G-G > A-G ~ G-T and for Ca^{2+} is G-G > A-G, with cations bound to the bases listed here.

No bending towards the minor groove?

The cation-induced bending discussed so far has been towards the major groove. To investigate the effect of cation binding within the minor groove, we examined the 28 Mg^{2+} and Ca^{2+} complexes listed in Table 1. Roll angles for 11 unique cases are plotted in Figure 11, with the high-resolution isomorphous C2 decamers on the left, and with bdj060, VTCa, and the lower-resolution $P2_12_12_1$ decamers on the right.

Consider first the bending that occurs across the TGG//CCA helix junction. As stated above for *CG*, *AG*, and CICa decamers, binding of Mg^{2+} and Ca^{2+} in the major and minor grooves at this junction results in a net bend of 15-20° from the second to fifth base. This bend can be seen in the normal vector plots of Figure 4, and its origin in roll angles is visible in parts A1 through A4 of Figure 11. Symbols along the bottom of each plot mark sites of minor groove divalent cation binding. But note that exactly the same roll bending across the helix

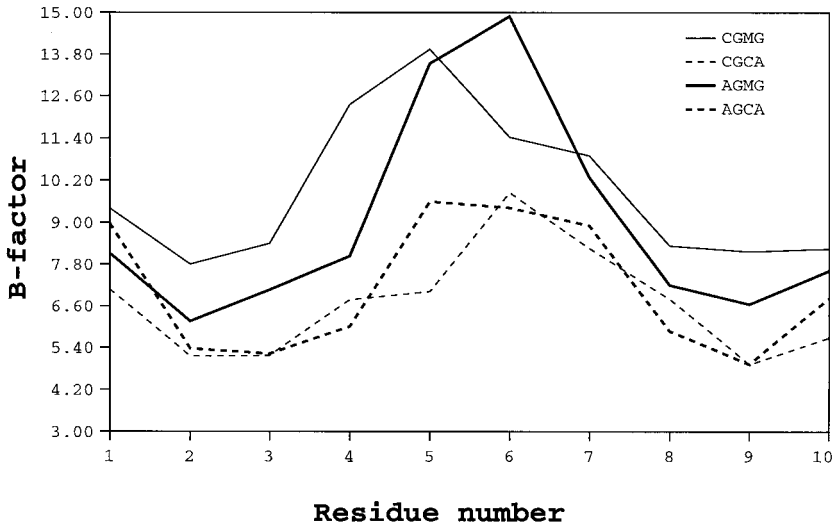

Figure 9. Plot of average *B*-value per DNA residue, with CGMg in thin continuous line, CGCa in thin broken line, AGMg in thick continuous line, and AGCa in thick broken line. Note that the average *B*-values of the calcium structures are lower than those of the magnesium structures, indicating calcium's greater stabilization of DNA than magnesium.

junction occurs in bdj008, although no minor groove cation is present. One could argue that the minor groove cation is present in bdj008, but is not observed because of the lower data resolution. However, this is unlikely to be the case, as the same authors who failed to observe a magnesium cation in bdj008 did in fact report Mg111 in bdj019 under similar experimental conditions (Table 1). Hence, this suggests that the minor groove cation Mg111 observed in *CG*, *AG*, and CICa decamers probably is not the cause of bending.

A second example of the non-effect of minor groove binding on DNA bending is the behavior of roll angle at the A3-A4 step. In this case, the roll angle remains strongly positive whether a minor groove cation is observed (part A3 of Figure 11) or not (A1 and A5). Very similar roll angles are observed in AGMg and AGCa, but a direct comparison with CGMg, CGCa, and bdj008 is biased both by the presence of a major groove cation at the A3-G4 step and by the difference in base sequence of residue 4.

A third example comes from comparison of the A-G steps in AGMg and AGCa (minor groove cations, large positive roll) with bdj008 (cations present, large negative roll). In all three cases a cation is observed in the minor groove, but the additional binding of a cation to the major groove in AGMg and AGCa converts the strongly negative roll angle (in bdj008) to positive. This strong reversal of roll might not be caused by the difference in cation binding, however, as the central two base-pairs in bdj008 have a G:A mismatch (G5:A16 and G15:A6).

The behavior of roll angles at the site of minor groove binding by divalent cations in the lower-resolution structures plotted on the right in Figure 11 ($\sim 0^\circ$ or slightly positive, but never strongly negative) is supportive of the idea that divalent cation binding in the minor groove has no effect on DNA bending. In all six instances, as discussed in Results, divalent cations in the minor groove in general sit squarely at the center of the groove and make cross-strand H-bonds with adjacent base and backbone atoms. In each $P2_12_12_1$ decamer (parts C1-C3 of Figure 11), only one minor groove cation is observed. It sits between base-pairs 4 and 5, with two of its water ligands each donating one hydrogen bond to imino or carboxyl atoms of residues 5 and 17 and one hydrogen bond to O4′ of residues 6 and 18. These interactions in effect prevent bending towards the minor groove. In bdj060 (part B of Figure 11), minor groove binding has no effect on bending, as the roll angles fluctuate around an average of 2°. Similarly, a minor groove Ca^{2+} outside of the A-tract causes no bending at all in bd0019 (part D of Figure 1).

In summary, the mechanics of *B*-DNA, coupled with the narrowness of the minor groove, allows the minor groove to open and close to accommodate a divalent cation with little or no bending towards the minor groove. But note that when the mechanics of *B*-DNA is changed, as by permanently attaching methyl groups to phosphate oxygen atoms, then bending in the direction of the minor groove becomes possible (Strauss-Soukup *et al.*, 1997). The rigidity of A-tracts and the extreme narrowness of its minor groove (2-4 Å), however, makes it difficult for the large bulky hydrated divalent cations ($\sim$5-7 Å) to bind to its minor groove. To our knowledge, such bulky cations have never been observed in the minor groove of A-tract structures in naked *B*-DNA.

Major groove bending at GGCC sequences and pyrimidine-purine steps outside A-tracts

The binding of divalent cations to the major groove of GGCC provides a structural explanation for many solution studies. The presence of significant numbers of phased A-tracts within anomalously migrating kinetoplast fragments at first suggested that the bending resided within A-tracts,

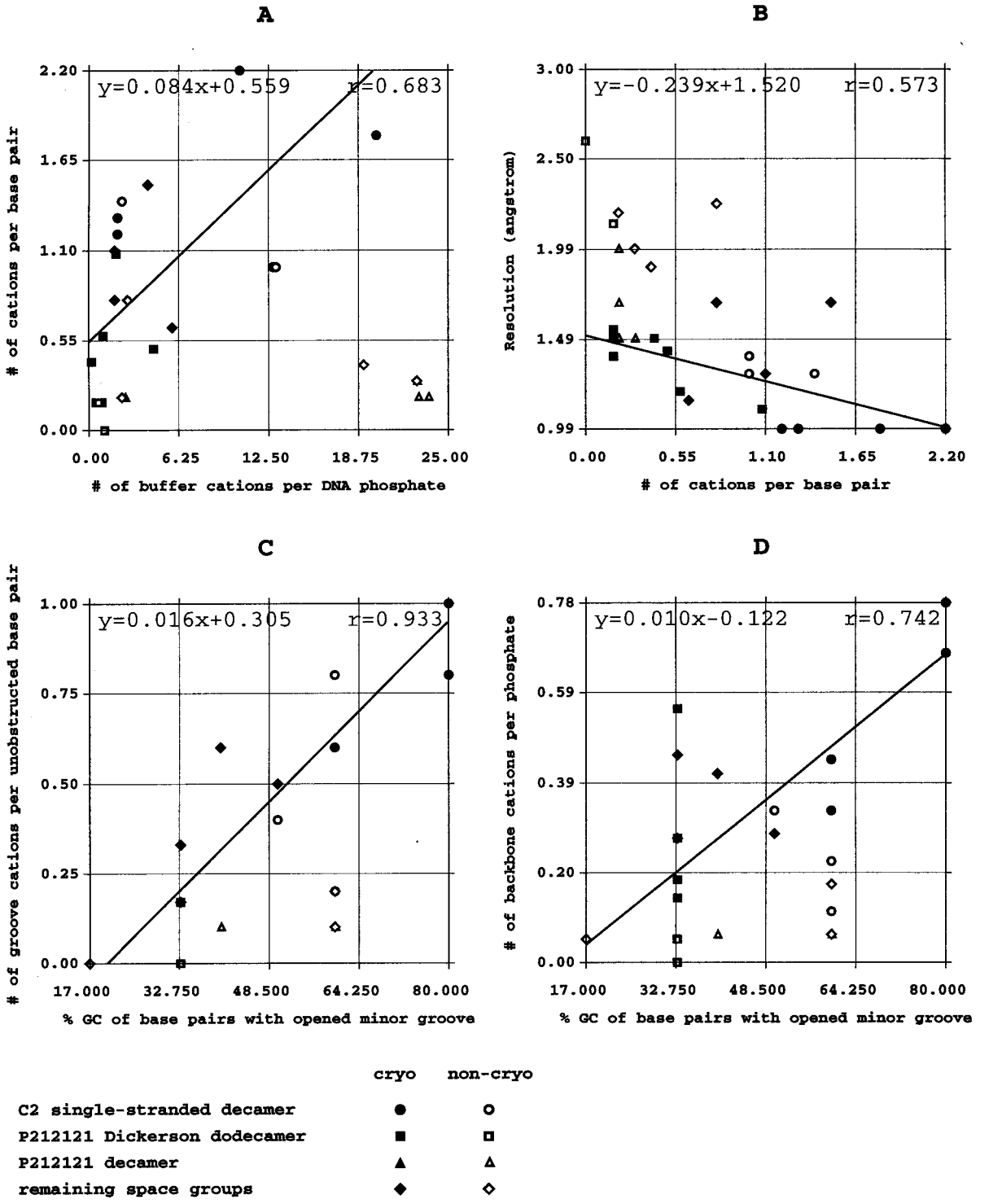

Figure 10. SHELXDNA correlation plots of helix variables: (a) cations per base-pair observed, *versus* cations per phosphate in the original crystallizing buffer, (b) resolution *versus* cations per base-pair, (c) groove cations per base-pair *versus* %GC for bases with unobstructed grooves, (d) backbone cations per phosphate, *versus* %GC as in (c). Circles, isomorphous monoclinic C2 decamers with one-half DNA duplex per asymmetric unit (first seven entries of Table 1). Triangles, decamers in space group $P2_12_12_1$. Squares, dodecamers CGCGAATTCGCG in space group $P2_12_12_1$. Diamonds, remaining structures. Filled symbols, data collection under cryogenic conditions ($\leqslant$100 K). Open symbols, data collected at higher temperature.

with bending in a direction that compressed the minor groove. However, later experiments indicated that cation-dependent bending in the direction of the major groove, with $Ca^{2+} > Mg^{2+} > Zn^{2+}$, resides within GGGCCC sequences instead (Brukner *et al.*, 1994). The same cation dependence

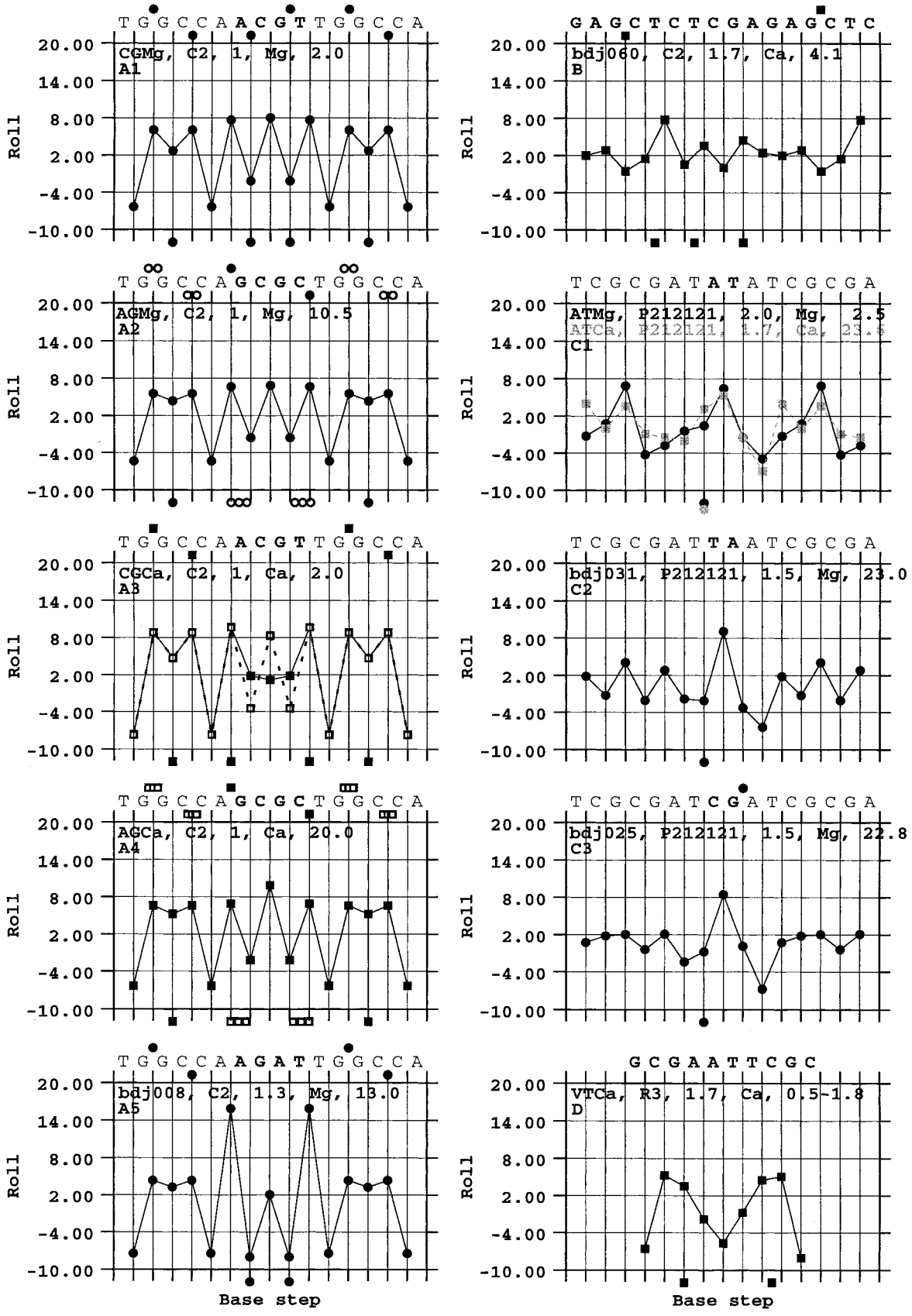

Figure 11 (*legend opposite*)

was observed in gel migration patterns of *Hin*dIII fragments (Li *et al*., 1998). Mg^{2+}-induced bending of another sort, by base-roll compression of the major groove at fragile pyrimidine-purine steps, has also been observed in 2.5-5 ns MDS of oligonucleotides 25-30 bp long (Young & Beveridge, 1998; Sprous *et al*., 1999). In these studies, A-tracts remained relatively straight, and Mg^{2+} were observed to be ''localized, in a hydrated form, in both grooves for periods exceeding 1 to 2 ns.''

Monovalent cations in higher order nucleic acid structures

The relative unimportance of monovalent cations compared to divalent cations on the structure of duplex DNA does not imply that monovalent cations are unimportant in other nucleic acids systems. Conversely, the importance of monovalent cations for the maintenance of these higher-ordered systems cannot be used as evidence for equivalent structural roles in duplex DNA. Nucleic acid structures beyond the double helix family (DNA triplex and quadruplex, tRNA, and RNA ribozyme) are structurally more complex molecules. Unlike the case for duplex DNA, where tertiary structure information determines where divalent cations bind, it is predominantly the quaternary structure requirement of these higher form molecules that determines where and which cations are needed. In quadruplexes, the close juxtaposition of eight guanine O6 atoms require the coordination of monovalent cations. The cation preference, $K^+ > Na^+ > Rb^+ > Cs^+ > Li^+$, is finely determined by size considerations (their corresponding ionic radii are 1.33, 0.95, 1.48, 1.69, and 0.6 Å; Rose & Hardin, 1994). Divalent cations are unfavorable for quadruplex formation because the extra positive charge would create destabilizing cation-cation repulsion. The bi-loop quadruplex pATTCATTC (Salisbury *et al*., 1997) and the AA-tetraloop of the P4-P6 domain of *Tetrahymena* ribozyme (Cate *et al*., 1996), provide further examples of the importance of monovalent cations in these systems. In the latter case, both chemical and geometric factors are involved in cation selection (Basu *et al*., 1998).

Conclusion

Monovalent and divalent cations are biologically important in nucleic acid structures. For duplex DNA, the relative lack of hydration of monovalent cations makes their interactions to base atoms largely electrostatic and relatively non-specific. Instead, they are ideal stabilization agents for higher-form nucleic acid structures, which usually have many negatively charged atoms in close proximity to one another. Divalent cations, on the other hand, tend to be fully hydrated. As such, their interactions to duplex DNA are more sequence-specific, because their water ligands can both donate and accept H-bonds to base atoms. They are drawn electrostatically to both the major and minor grooves of DNA, as well as to the sugar-phosphate backbone. Once close to DNA, sequence-specific interactions with base atoms determine where they bind. In the minor groove, divalent cations almost exclusively form cross-strand H-bonds with base atoms of adjacent base-pairs and with sugar O4′ atoms of the opposing groove walls. These interactions, coupled with the narrowness of the minor groove, prevent bending towards the minor groove. In the much wider major groove, divalent cations tend to interact with base atoms of adjacent atoms on the purine strand. At GGCC/GGCC sequences, binding of magnesium or calcium to the major groove GG residues, and to the minor groove at the GC step, result in bending of the helix towards the major groove. For the first time, polar covalent bonds between Ca^{2+} and adjacent O6 and N7 atoms of stacked Gs are observed. In addition to having higher affinity to DNA, Ca^{2+} binding to DNA also results in greater DNA bending and thermal stabilization than Mg^{2+}.

Materials and Methods

Oligonucleotide synthesis and purification

The DNA decamers were synthesized at the 10 μmole scale on an Eppendorf ECOSYN D300 synthesizer. After deprotection in concentrated NH_4OH at 55 °C for 16 hours, they were dried and purified on a DE52 column with a 50-350 mM KCl gradient in 50 mM Tris (pH 7.5), 7 M urea. Peak fractions were pooled and loaded on a small DE52 column. Tris, KCl, and urea were removed by washing with 50 mM triethylamine bicarbonate (TEAB) and the DNA eluted with a solution of 0.85 M TEAB and 15% (v/v) ethanol. TEAB and ethanol were removed by repeated drying on a Rotovap and the DNA lyophilized to dryness, resuspended in double distilled water, and stored at −80 °C until crystallization.

Figure 11. Roll angles and divalent cation loci for isomorphous C2 decamers (A1-A5), bdj060 (B), isomorphous $P2_12_12_1$ decamers (C1-C3), and VTCa (D). The DNA sequence is given at the top of each plot. Bases which differ within each isomorphous are in bold type. Mg^{2+} are represented by circles and Ca^{2+} are represented by squares. Filled symbols represent atoms in single conformations, and opened symbols represent atoms in multiple conformations. Symbols along the top edge of each plot indicate major groove cation binding, and along the bottom edge, minor groove binding. For the major groove, where binding can be to one side of the groove or the other, symbols above the base sequence represent binding to that sequence, whereas symbols below the sequence represent binding to the other strand (not shown). Stacked helices are drawn as though their repeating sequences were continuous. For all but VTCA (D), a decamer commences with the fourth base as the sequence is drawn.

Crystallization and data collection

Crystals were grown by sitting drop vapor diffusion at 4 °C in nine-well plates. Crystals could be grown with $MgCl_2$, MgOAc, $CaCl_2$, or CaOAc, and at various salt concentrations. Salt concentrations were chosen to counter-balance the 36 DNA backbone phosphate oxygen atoms. For consistency the crystals chosen for data collection all have acetate as their counter-anion (Table 2A). Survey precession photos showed these to be isomorphous with the previously published CCAACGTTGG decamer (Privé *et al*., 1991). Oscillation data up to 1.35 Å (AGMg) were collected initially on a local rotating anode X-ray source with a Rigaku Raxis-II image detector at 100 K. Data processing and structure solution were carried out exactly as described previously (Chiu *et al*., 1999). Atomic-resolution data sets were recollected at NSLS beamline X8C on a Mar 30 cm detector (Table 2A). All subsequent discussion pertains to these synchrotron data sets.

Structure determination of atomic resolution data sets

Each structure was solved using the previously determined lower-resolution structures, with all solvent atoms removed, as the starting model. 5% of the reflections in each data set were randomly set aside for structure validation by the Free-*R* test. The first seven cycles of refinement were done in XPLOR. Subsequent refinement based on intensities, with the same reflections flagged for the Free-*R* test as in XPLOR, were performed using *Shelxl* (Sheldrick & Schneider, 1997). Discrete clusters of excess density in $F_o - F_c$ maps contoured 3.5σ next to a DNA backbone fragment that has been modelled in one conformation were treated as follows: the occupancy of the major conformation was refined first, then the second or third conformations were assigned and refined based on its $F_o - F_c$ map, and finally atoms of all conformations were refined with individual ADPs. SIMU and ISOR restraints were optimized using the Free-*R*. Hydrogen atoms with ideal geometry were modeled based on the coordinates of heavy atoms they are attached to. Refinement statistics are given in Table 2B.

Identification of Mg^{2+} and Ca^{2+} complexes

Clusters of closely-spaced electron density peaks in $2F_o - F_c$ and $F_o - F_c$ maps were assigned as hydrated cation complexes only if they satisfy all three criteria: (1) they have proper coordination geometry and uniform bond length, (2) the $F_o - F_c$ difference maps before and after peak assignment are not abnormal, and (3) the *B*-values of the assigned atoms are uniformly equivalent. When in doubt, a cluster was not assigned until the model had improved enough to allow a more reliable assessment of the three criteria.

A cluster of up to six peaks arranged in a regular octahedral geometry ca. 2.1 Å around the central peak is the hallmark signature of a Mg^{2+} complex. A cluster of up to six or seven peaks, arranged in an irregular geometry ca. 2.45 Å around the central peak can either be a Na^+ or Ca^{2+} complex, respectively. At atomic resolution, the ambiguity can be resolved by examining relative peak heights of the central atom *versus* those of the surrounding atoms. Ca^{2+}, with 18 electrons compared to the ten of Na^+, diffracts more strongly than Na^+. Hence, in the refined $F_o - F_c$ map a calcium complex that had been tentatively assigned incorrectly as a Na^+ complex would exhibit a positive overlapping of the central atom. Conversely, a Na^+ complex that had been assigned erroneously as a Ca^{2+} complex would have a negative peak overlapping the central atom in the refined $F_o - F_c$ map.

Water molecules in the coordinating sphere of Mg^{2+} were restrained to an octahedral geometry by imposing bond length (DFIX 2.08 MG OW1...MG OW6) and bond angle (DANG 4.16 OW1 OW2 OW3 OW4 OW5 OW6, DANG 2.94 OW1 OW3...OW4 OW6) restraints, where water molecules OW1 and OW2, OW3 and OW4, and OW5 and OW6 sit across each other. Because of the irregular geometry around calcium, only bond length restraints (DFIX 2.45 CA OW1...CA OW7) were imposed. If both bond lengths and *B*-values varied widely upon refinement, the assignment of an atom cluster as a cation complex was deleted. In cases where an $F_o - F_c$ map revealed multiple conformations for the cation, each conformation and its occupancy were refined individually, and then the multiple conformations were combined, with the sum of their occupancies restrained to 1.00. In cases where the disordered cation was located near a disordered DNA residue, its occupancy was coupled to that of the disordered DNA component.

Using the above criteria, water residue 30 in bdj052 (Grzeskowiak *et al*., 1993), which sits in the center of a planar quadrilateral arrangement of atoms consisting of symmetry-related atoms 17 O2P and 72 OW, each 2.46 Å away, has been re-assigned as a calcium atom.

Shelxdna

The program SHELXDNA was written to facilitate refinement and analysis of X-ray crystal structures by *Shelxl-97,* to plot and analyze one or more CURVES and FREEHELIX output files, and can be obtained by email from thang@mbi.ucla.edu or at www.doe-mbi.ucla.edu/people/thang/thang.html. When the program is run, it presents a list of options to the user, from which the user enters an integer, or a carriage return to exit. In general, default values are accepted by entering carriage return. The first six options are to help with refinement in SHELXL-97. When going from an XPLOR or CNS refinement to SHELXL-97, one has to carry over the Free-*R* flags. SHELXPRO does this for structure amplitudes, but not for structure intensities. Hence, SHELXPRO is first used to convert an XPLOR/CNS amplitude file, which could have been modified in CNS or CCP4, to SHELXL-97 format; this carries over the Free-*R* flags. SHELXPRO is then used to convert the original unmodified SCALEPACK intensity file to SHELXL-97 format, and option 1 of SHELXDNA takes these two files and replaces the Free-*R* flags in the intensity file with those from the amplitude file.

The number of cycles needed to reach convergence can be determined graphically by plotting the weighted *R*2 *versus* cycle number. The standard output from SHELXL-97 is first redirected to a file and one or more of these files are input into the second option of SHELXDNA. When more then one file is entered, the entire progress of refinement over the course of model building and refinement can be plotted together. After refinement in SHELXL-97 has proceeded, SHELXPRO is used to generate an instruction file for the next cycle. Unfortunately, it does not generate dictionary parameters for nucleic acids. These can be restored from a previous instruction file by using a text editor, or by using the third option of SHELXDNA. The *B*-values and individual anisotropic displacement parameters (ADPs)

can also be plotted by SHELXDNA. When *B*-values are plotted, a .pdb file from any refinement package is required, and when ADPs are plotted, both .pdb and .lst files from SHELXL-97 are required. These two options plot *B*-values or ADPs for each atom and each residue, and also print them out along with average values per atom-type and per molecule. The number of x and y intervals, as well as the minimal and maximal values can be specified by the user or automatically determined by the program. During model-building in ONO, if one wants to write out a coordinate file from a symmetry object, all information on each atom's occupancy, temperature factor, and ADPs are lost. To restore these, the original coordinate used to generate the symmetry objects in ONO, and the coordinate generated by ONO, are entered into SHELXDNA.

At any stage of the refinement, the base-pair and base step parameters can be calculated using either the program FREEHELIX or CURVES. These parameters can be conveniently plotted automatically by SHELXDNA. At present, only parameters for duplex DNA can be plotted; parameters for triplex or quadruplex are not handled. When these plotting options are chosen, the user is first asked to enter an output file name and the filename which contains the standard output from FREEHELIX or CURVES. When more than one filename is given (in accordance to the program's prompt), a whole database of structures can be plotted simultaneously. When more than one file is plotted, the x-axis of each DNA fragment can be offset by an integral value relative to another. Next, the user is asked to choose customized or default axes limits. If the customized option is chosen, the user is presented with a list of parameters to set the limits. Next, the user is prompted for inputs which affect the appearance of the plots. These includes options for setting linewidth, linedash, symbol, and color, as well as labelling and connecting of data points. The user can either enter a value in response to each prompt, or enter a carriage return to accept default values generated by the program. Finally, the user is presented with a list of FREEHELIX or CURVES parameters to plot. The user can plot either one or more of the listed parameters, all parameters, or one parameter *versus* another. Plotting all parameters at once will give six plots per page whereas plotting one parameter at a time will give one plot per page. Figures 35 and 89 were plotted with SHELXDNA. In Figure 10, correlation values and best-fit lines were determined by Cricket Graph.

Data Base Accession Number

The crystal structures of the Mg^{2+} and Ca^{2+} salts of CCAACGTTGG and CCAGCGCTGG have been assigned PDB accession numbers 1en3, 1en1, 1en9 and 1ene, respectively or NDB accession numbers bd0033, bd0034, bd0035 and bd0036.

Acknowledgments

We thank Mary L. Kopka for help with DNA purification and crystallization, Dr David S. Goodsell for help with data collection and structure determination of the low resolution data sets, Dr Duilio Cascio for help with synchrotron data collection, and Dr Duilio Cascio and Rick Farhner for help with *Shelxl* and *Ono*. This work is supported by NIH grant GM-31299 and a Dissertation Year Fellowship from UCLA Graduate Division to T.K.C.

References

Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J. R. & Dickerson, R. E. (1993). The crystal structure of the trigonal decamer CGATCG-A^{6ME}-TCG: a *B*-DNA helix with 10.6 base-pairs per turn. *J. Mol. Biol.* **231**, 768-784.

Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferre-D Amare, A. R., Strobel, S. A. & Doudna, J. A. (1998). A specific monovalent metal ion integral to the AA platform of the RNA tetraloop receptor. *Nature Struct. Biol.* **5.11**, 986-992.

Behe, M. & Felsenfeld, G. (1981). Effects of methylation on a synthetic polynucleotide: the B-Z transition in poly(dG-m5dC)·poly(dG-m5dC). *Proc. Natl Acad. Sci. USA,* **78**, 1619-1623.

Berger, I., Tereshko, V., Ikeda, H., Marquez, V. E. & Egli, M. (1998). Crystal structures of *B*-DNA with incorporated 2 2′-deoxy-2′-fluoro-arabino- furanosyl thymines: implications of conformational preorganization for duplex stability. *Nucl. Acids Res.* **26**, 2473.

Boynton, A. L., McKeehan, W. L. & Whitfield, J. F. (1982). *Ions, Cell Proliferation and Cancer*, Academic Press, NY.

Braunlin, W. H. (1995). NMR studies of cation-binding environments on nucleic acids. In *Advan. Biophys. Chem.* (Bush, C. A., ed.), vol. 5, pp. 89-139, JAI Press, Greenwich.

Brukner, I. S., Susic, M., Dlakic, A. S. & Pongor, S. (1994). Physiological concentration of Mg^{2+} ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *J. Mol. Biol.* **236**, 26-32.

Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, R. R. & Doudna, J. A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science,* **273**, 1678-1685.

Chiu, T. K., Kaczor-Grzeskowiak, M. & Dickerson, R. E. (1999). Absence of minor groove monovalent cations in the crosslinked dodecamer CGCGA-ATTCGCG. *J. Mol. Biol.* **292**, 589-608.

Collins, K. D. (1997). Charge density-dependent strength of hydration and biological structure. *Biophys. J.* **8**, 1426-1433.

Denisov, V. P. & Halle, B. (2000). Sequence-specific binding of counterions to *B*-DNA. *Proc. Natl Acad. Sci. USA,* **97**, 629-633.

Dickerson, R. E. (1998). DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucl. Acids Res.* **26.8**, 1906-1926.

Dickerson, R. E. & Chiu, T. K. (1997). Helix bending as a factor in protein/DNA recognition. *Biopolymers Nucl. Acid Sci.* **44**, 361-403.

Diekmann, S. & Wang, J. C. (1985). On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mobilities. *J. Mol. Biol.* **186**, 1-11.

Draper, D. E. & Misra, V. K. (1998). RNA shows its metal. *Nature Struct. Biol.* **927-930**, 5.1.

Drew, H. R. & Dickerson, R. E. (1981). Structure of a *B*-DNA dodecamer: III. Geometry of hydration. *J. Mol. Biol.* **151**, 535-556.

Eichhorn, G. L. & Shin, Y. A. (1968). Interactions of metal ions with polynucleotides and related com-

pounds. XII. The relative effect of various metal ions on DNA helicity. *J. Am. Chem. Soc.* **90**, 7323-7328.

El Hassan, M. A. & Calladine, C. R. (1995). The assessment of the geometry of dinucleotide steps in double-helical DNA: a new local calculation scheme. *J. Mol. Biol.* **251**, 648-664.

Feig, M. & Pettitt, B. M. (1999). Sodium and chlorine ions as part of the DNA solvation shell. *Biophys. J.* **77**, 1769-1781.

Goodsell, D. S., Kopka, M. L., Cascio, D. & Dickerson, R. E. (1993). Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl Acad. Sci. USA,* **90**, 2930-2934.

Goodsell, D. S., Grzeskowiak, K. & Dickerson, R. E. (1995). Crystal structure of CTCTCGAGAG: implications for the structure of the Holliday junction. *Biochemsitry,* **34**, 1022.

Grzeskowiak, K., Yanagi, K., Privé, G. G. & Dickerson, R. E. (1991). The structure of B-helical CGATCGATCG and comparison with CCAACGTTGG. The effect of base-pair reversals. *J. Biol. Chem.* **266**, 8861-8883.

Grzeskowiak, K., Goodsell, D. S., Kaczor-Grzeskowiak, M., Cascio, D. & Dickerson, R. E. (1993). Crystallographic analysis of CCAAGCTTGG and its implications for bending in *B*-DNA. *Biochemistry,* **32**, 8923-8931.

Hunter, C. A. & Lu, X.-J. (1997). DNA base-stacking interactions: a comparison of theoretical calculations with oligonucleotides X-ray crystal structures. *J. Mol. Biol.* **265**, 603-619.

Juo, S. Z., Chiu, T. K., Leiberman, P. M., Baikalov, I., Berk, A. J. & Dickerson, R. E. (1996). How proteins recognize the TATA Box. *J. Mol. Biol.* **261**, 239-254.

Kopka, M. L., Han, G. W., Goodsell, D. S., Chiu, T. K., Walker, W. L., Lown, J. W. & Dickerson, R. E. (1998). DNA sequence recognition in the minor groove by polyamides, using a GC-specific reading element: a perspective from crystallography. In *Structure, Motion, Interaction and Expression of Biological Macromolecules*, pp. 177-191, Proceedings of the Tenth Conversation, State University of New York, Albany, NY.

Korolev, N., Lyubartsev, A. P., Rupprecht, A. & Nordenskiold, L. (1999). Competitive binding of Mg^{2+}, Ca^{2+}, Na^{+}, and K^{+} ions to DNA in oriented DNA fibers: experimental and monte carlo simulation results. *Biophys. J.* **77**, 2736-2749.

Langlais, M., Tajmir-Riahi, H. A. & Savoie, R. (1990). Raman spectroscopic study of the effects of Ca^{2+}, Mg^{2+}, Zn^{2+} and Cd^{2+} ions on calf thymus DNA: binding sites and conformational changes. *Biopolymers,* **30**, 743-752.

Lavery, R. & Sklenar, H. (1988). The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63-91.

Leonard, G. A. & Hunter, W. N. (1993). Crystal and molecular structure of d(CGTAGATCTACG) at 2.25 Å resolution. *J. Mol. Biol.* **234**, 198.

Li, A. Z., Huang, H., Re, X., Qi, L. J. & Marx, K. A. (1998). A gel electrophoresis study of the competitive effects of monovalent counterion on the extent of divalent counterions binding to DNA. *Biophys. J.* **74**, 964-973.

Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J. & Dickerson, R. E. (1993). The structure of the *B*-DNA decamer CCAACITTGG in two different space groups: conformational flexibility of *B*-DNA. *Biochemistry,* **32**, 1373-1389.

Liu, J., Malinina, L., Huynh-Dinh, T. & Subirina, J. A. (1998). The structure of the most studied DNA fragment changes under the influence of ions:a new packing of d(CGCGAATTCGCG). *FEBS Letters,* **438**, 211-214.

Minasov, G., Tereshko, V. & Egli, M. (1999). Atomic-resolution crystal structures of *B*-DNA reveal specific influences of divalent metal ions on conformation and packing. *J. Mol. Biol.* **291**, 83-99.

Narasimhan, V. & Bryan, A. M. (1984). Conformational flexibility of poly(dG-m5dC) under very low salt conditions. *Experientia,* **40**, 827-828.

Neault, J. F. & Tajmir-Riahi, H. A. (1999). Structural analysis of DNA-chlorophyll complexes by fourier transform infrared difference spectroscopy. *Biophys. J.* **76**, 2177-2182.

Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T. & Berman, H. (1996). New parameters for the refinement of nmucleic acid containing structures. *Acta Crystallog. sect. D,* **52**, 57-64.

Portmann, S., Altman, K.-H., Reynes, N. & Egli, M. (1996). Crystal structures of oligodeoxyribonucleotides containing 6′-α-methyl and 6′-α-hydroxy carbocyclic thymidines. *J. Am. Chem. Soc.* **119**, 2396-2403.

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L. & Dickerson, R. E. (1987). Helix geometry, hydration, and G-A mismatch in a *B*-DNA decamer. *Science,* **238**, 498-504.

Privé, G. G., Yanagi, K. & Dickerson, R. E. (1991). Structure of the *B*-DNA decamer CCAACGTTGG and comparison with isomorphous decamers CCAAGATTGG and CCAGGCCTGG. *J. Mol. Biol.* **217**, 177-199.

Pullman, A. & Pullman, B. (1981). Molecular electrostatic potential of the nucleic acids. *Quart. Rev. Biophys.* **14**, 285.

Quintana, J. R., Grzeskowiak, K., Yanagi, K. & Dickerson, R. E. (1992). The structure of a *B*-DNA decamer with a central T-A step: CGATTAATCG. *J. Mol. Biol.* **225**, 379-395.

Rose, W. S. & Hardin, C. C. (1994). Ion-induced stabilization of the G-DNA quadruplex: free energy perturbation studies. *J. Am. Chem. Soc.* **111**, 6070-6080.

Salisbury, S. A., Wilson, S. E., Powell, H. R., Kennard, O., Lubini, P., Sheldrick, G. M., Escaja, N., Alazzouzi, E. M., Granda, A. & Pedroso, E. (1997). The bi-loop, a new general four-stranded DNA motif. *Proc. Natl Acad. Sci. USA,* **94**, 5515-5518.

Shatzky-Schwartz, M., Arbuckle, N. D., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E., Luisi, B. F. & Shakked, Z. (1997). X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract dependent curvature. *J. Mol. Biol.* **267**, 595-623.

Sheldrick, G. M. & Schneider, T. R. (1997). SHELX-97: high resolution refinement. *Methods Enzymol.* **276**, 319-343.

Sprous, D., Young, M. A. & Beveridge, D. L. (1999). Molecular dynamics studies of axis bending in $d(G_5\text{-}(GA_4T_4)_2\text{-}C_5)$ and $d(G_5\text{-}(GT_4A_4C)_2\text{-}C_5)$: effects of sequence polarity on DNA curvature. *J. Mol. Biol.* **285**, 1623-1632.

Shui, X., McFail-Isom, L., Hu, G. G. & Williams, L. D. (1998a). The *B*-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry,* **37**, 8341-8355.

Shui, X., Sines, C. C., McFail-Isom, L., VanDerveer, D. & Williams, L. D. (1998b). Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry*, **37**, 16877-16887.

Strauss-Soukup, J. K., Vaghefi, M. M., Hogrefe, R. I. & Maher, L. J., III (1997). Effects of neutralization pattern and stereochemistry on DNA bending by methylphosphonate substitutions. *Biochemistry,* **36**, 8692-8698.

Tereshko, V., Minasov, G. & Egli, M. (1999a). The DICKERSON-DREW *B*-DNA dodecamer revisited - at atomic resolution. *J. Am. Chem. Soc.* **121**, 470-471.

Tereshko, V., Minasov, G. & Egli, M. (1999b). A ''Hydrat-ion spine'' in a *B*-DNA minor groove. *J. Am. Chem. Soc.* **121**, 3590-3595.

van de Sande, J. H., McIntosh, L. P. & Jovin, T. M. (1982). Mn^{2+} and other transition metals at low concentration induce the right-to-left helical transformation of poly[d(G-C)]. *EMBO J.* **1**, 777-782.

Vlieghe, D., Turkenburg, J. P. & Meervelt, L. V. (1999). *B*-DNA at atomic resolution reveals extended hydration patterns. *Acta Crystallog. sect. D,* **55**, 1495-1502.

Young, M. A. & Beveridge, D. L. (1998). Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.* **281**, 675-687.

Young, M. A., Ravishanker, G., Beveridge, D. L. & Berman, H. (1995). Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophys. J.* **68**, 2454-2468.

Young, M. A., Jayaram, B. & Beveridge, D. L. (1997). Intrusion of counterions into the spine of hydration in the minor groove of *B*-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* **119**, 59-69.

Yuan, H., Quintana, J. R. & Dickerson, R. E. (1992). Alternative structures for alternating poly(dA-dT) tracts: the structure of the *B*-DNA decamer CGATATATCG. *Biochemistry*, **31**, 8009-8021.

Zacharias, W., Larson, J. E., Klysik, J., Stirdivant, S. M. & Wells, R. D. (1982). Conditions which cause the right-handed to left-handed DNA conformational transitions. Evidence for several types of left-handed DNA structures in solution. *J. Biol. Chem.* **257**, 2775-2782.

Edited by I. Tinoco

(Received 27 March 2000; received in revised form 23 June 2000; accepted 30 June 2000)